Lecture 5 - UM 102 (Spring 2021)

Vamsi Pritham Pingali

IISc

Recap

Recap

- Defined orthogonality and proved that non-zero orthogonal elements are linearly independent.

Recap

- Defined orthogonality and proved that non-zero orthogonal elements are linearly independent.
- Proved Parseval's formula.

Recap

- Defined orthogonality and proved that non-zero orthogonal elements are linearly independent.
- Proved Parseval's formula.
- Gram-Schmidt orthogonalisation procedure.

Orthogonal complement

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space.

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S.

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW).

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.
- Examples :

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.
- Examples :
  - The perpendicular subspace to the set $\{(1,1),(1,2)\}$ in $\mathbb{R}^2$ with the usual inner product is

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.
- Examples :
  - The perpendicular subspace to the set $\{(1,1),(1,2)\}$ in $\mathbb{R}^2$ with the usual inner product is $\{(0,0)\}$.

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.
- Examples :
  - The perpendicular subspace to the set $\{(1,1),(1,2)\}$ in $\mathbb{R}^2$ with the usual inner product is $\{(0,0)\}$. Indeed, $(a,b).(1,1) = a+b = 0$ and $(a,b).(1,2) = a+2b = 0$ imply that $a = b = 0$.

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.
- Examples :
  - The perpendicular subspace to the set $\{(1,1),(1,2)\}$ in $\mathbb{R}^2$ with the usual inner product is $\{(0,0)\}$. Indeed, $(a,b).(1,1) = a+b = 0$ and $(a,b).(1,2) = a+2b = 0$ imply that $a = b = 0$.
  - Given a line $t(1,2,3)$ in $\mathbb{R}^3$,

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.
- Examples :
    - The perpendicular subspace to the set $\{(1,1),(1,2)\}$ in $\mathbb{R}^2$ with the usual inner product is $\{(0,0)\}$. Indeed, $(a,b).(1,1) = a+b = 0$ and $(a,b).(1,2) = a+2b = 0$ imply that $a = b = 0$.
    - Given a line $t(1,2,3)$ in $\mathbb{R}^3$, its orthogonal complement is a plane:

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.
- Examples :
  - The perpendicular subspace to the set $\{(1,1),(1,2)\}$ in $\mathbb{R}^2$ with the usual inner product is $\{(0,0)\}$. Indeed, $(a,b).(1,1) = a+b = 0$ and $(a,b).(1,2) = a+2b = 0$ imply that $a = b = 0$.
  - Given a line $t(1,2,3)$ in $\mathbb{R}^3$, its orthogonal complement is a plane: $0 = (x,y,z).(1,2,3) = x+2y+3z$.

Orthogonal complement

- Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is denoted as $S^{\perp}$.
- $S^{\perp}$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^{\perp}$ is called the *orthogonal complement* of S.
- Examples :
  - The perpendicular subspace to the set $\{(1,1),(1,2)\}$ in $\mathbb{R}^2$ with the usual inner product is $\{(0,0)\}$. Indeed, $(a,b).(1,1) = a+b = 0$ and $(a,b).(1,2) = a+2b = 0$ imply that $a = b = 0$.
  - Given a line $t(1,2,3)$ in $\mathbb{R}^3$, its orthogonal complement is a plane: $0 = (x,y,z).(1,2,3) = x+2y+3z$.
  - The continuous functions orthogonal to 1 with the integration inner product on $[0,1]$ are the ones with zero average.

Orthogonal decomposition

Orthogonal decomposition

- Recall how we used $(1,1)$ and $(1,2)$ to create orthogonal vectors.

Orthogonal decomposition

- Recall how we used $(1,1)$ and $(1,2)$ to create orthogonal vectors.
- Motivated by this construction,

Orthogonal decomposition

- Recall how we used $(1, 1)$ and $(1, 2)$ to create orthogonal vectors.
- Motivated by this construction, we have a theorem: Let $(V, \langle , \rangle)$ be an inner product space and $S \subseteq V$ be a f.d. subspace.

Orthogonal decomposition

- Recall how we used $(1,1)$ and $(1,2)$ to create orthogonal vectors.
- Motivated by this construction, we have a theorem: Let $(V, \langle , \rangle)$ be an inner product space and $S \subseteq V$ be a f.d. subspace. Then every element $x \in V$ can be represented uniquely as a sum $x = s + s^{\perp}$ where $s \in S$ and $s^{\perp} \in S^{\perp}$.

Orthogonal decomposition

- Recall how we used $(1,1)$ and $(1,2)$ to create orthogonal vectors.
- Motivated by this construction, we have a theorem: Let $(V, \langle , \rangle)$ be an inner product space and $S \subseteq V$ be a f.d. subspace. Then every element $x \in V$ can be represented uniquely as a sum $x = s + s^{\perp}$ where $s \in S$ and $s^{\perp} \in S^{\perp}$. Moreover, $\|x\|^2 = \|s\|^2 + \|s^{\perp}\|^2$.

Orthogonal decomposition

- Recall how we used $(1, 1)$ and $(1, 2)$ to create orthogonal vectors.
- Motivated by this construction, we have a theorem: Let $(V, \langle , \rangle)$ be an inner product space and $S \subseteq V$ be a f.d. subspace. Then every element $x \in V$ can be represented uniquely as a sum $x = s + s^{\perp}$ where $s \in S$ and $s^{\perp} \in S^{\perp}$. Moreover, $\|x\|^2 = \|s\|^2 + \|s^{\perp}\|^2$.
- Caveat: If S is not f.d., the above result is NOT true in general!

Orthogonal decomposition

Orthogonal decomposition

- Proof:

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$.
  $\langle x - s, s_1 \rangle = \sum_j \bar{c}_j \langle x, e_j \rangle - \sum_{j,k} \bar{c}_j \langle x, e_i \rangle \langle e_i, e_j \rangle$.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$.
  $\langle x - s, s_1 \rangle = \sum_j \bar{c}_j \langle x, e_j \rangle - \sum_{j,k} \bar{c}_j \langle x, e_i \rangle \langle e_i, e_j \rangle$. By orthonormality, the latter is $\sum_j \bar{c}_j \langle x, e_j \rangle - \sum_j \bar{c}_j \langle x, e_j \rangle = 0$.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$. $\langle x - s, s_1 \rangle = \sum_j \bar{c}_j \langle x, e_j \rangle - \sum_{j,k} \bar{c}_j \langle x, e_i \rangle \langle e_i, e_j \rangle$. By orthonormality, the latter is $\sum_j \bar{c}_j \langle x, e_j \rangle - \sum_j \bar{c}_j \langle x, e_j \rangle = 0$. Hence, $s^{\perp} = x - s \in S^{\perp}$.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$.
$\langle x - s, s_1 \rangle = \sum_j \bar{c}_j \langle x, e_j \rangle - \sum_{j,k} \bar{c}_j \langle x, e_i \rangle \langle e_i, e_j \rangle$. By orthonormality, the latter is $\sum_j \bar{c}_j \langle x, e_j \rangle - \sum_j \bar{c}_j \langle x, e_j \rangle = 0$.
Hence, $s^{\perp} = x - s \in S^{\perp}$. If $x = t + t^{\perp} = s + s^{\perp}$, then $t - s = s^{\perp} - t^{\perp} \in S \cap S^{\perp} = \{0\}$.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$.
  $\langle x - s, s_1 \rangle = \sum_j \bar{c}_j \langle x, e_j \rangle - \sum_{j,k} \bar{c}_j \langle x, e_i \rangle \langle e_i, e_j \rangle$. By orthonormality, the latter is $\sum_j \bar{c}_j \langle x, e_j \rangle - \sum_j \bar{c}_j \langle x, e_j \rangle = 0$. Hence, $s^{\perp} = x - s \in S^{\perp}$. If $x = t + t^{\perp} = s + s^{\perp}$, then $t - s = s^{\perp} - t^{\perp} \in S \cap S^{\perp} = \{0\}$.
  $\|x\|^2 = \|s\|^2 + \|s^{\perp}\|^2 + \langle s, s^{\perp} \rangle + \langle s^{\perp}, s \rangle$ but $\langle s, s^{\perp} \rangle = 0$.

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$.
  $\langle x - s, s_1 \rangle = \sum_j \bar{c}_j \langle x, e_j \rangle - \sum_{j,k} \bar{c}_j \langle x, e_i \rangle \langle e_i, e_j \rangle$. By orthonormality, the latter is $\sum_j \bar{c}_j \langle x, e_j \rangle - \sum_j \bar{c}_j \langle x, e_j \rangle = 0$.
  Hence, $s^{\perp} = x - s \in S^{\perp}$. If $x = t + t^{\perp} = s + s^{\perp}$, then $t - s = s^{\perp} - t^{\perp} \in S \cap S^{\perp} = \{0\}$.
  $\|x\|^2 = \|s\|^2 + \|s^{\perp}\|^2 + \langle s, s^{\perp} \rangle + \langle s^{\perp}, s \rangle$ but $\langle s, s^{\perp} \rangle = 0$.
- The element $s = \sum_i \langle x, e_i \rangle e_i$ is called the *orthogonal projection* of x on the (f.d.) subspace S.

Orthogonal decomposition

- Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$.
  $\langle x - s, s_1 \rangle = \sum_j \bar{c}_j \langle x, e_j \rangle - \sum_{j,k} \bar{c}_j \langle x, e_i \rangle \langle e_i, e_j \rangle$. By orthonormality, the latter is $\sum_j \bar{c}_j \langle x, e_j \rangle - \sum_j \bar{c}_j \langle x, e_j \rangle = 0$.
  Hence, $s^{\perp} = x - s \in S^{\perp}$. If $x = t + t^{\perp} = s + s^{\perp}$, then $t - s = s^{\perp} - t^{\perp} \in S \cap S^{\perp} = \{0\}$.
  $\|x\|^2 = \|s\|^2 + \|s^{\perp}\|^2 + \langle s, s^{\perp} \rangle + \langle s^{\perp}, s \rangle$ but $\langle s, s^{\perp} \rangle = 0$.
- The element $s = \sum_i \langle x, e_i \rangle e_i$ is called the *orthogonal projection* of x on the (f.d.) subspace S. It is basically the "shadow" of x on S.

The approximation problem

The approximation problem

- Consider the following questions:

The approximation problem

- Consider the following questions:
  - What is the best way to approximate continuous functions using sines and cosines ?

- Consider the following questions:
  - What is the best way to approximate continuous functions using sines and cosines ?
  - What is the best way to approximate continuous functions using polynomials ?

The approximation problem

- Consider the following questions:
  - What is the best way to approximate continuous functions using sines and cosines ?
  - What is the best way to approximate continuous functions using polynomials ?
  - If we plot the price of houses vs their area (in a particular locality)

The approximation problem

- Consider the following questions:
  - What is the best way to approximate continuous functions using sines and cosines ?
  - What is the best way to approximate continuous functions using polynomials ?
  - If we plot the price of houses vs their area (in a particular locality) what is the "best" estimate of price per square foot ?

The approximation problem

- Consider the following questions:
  - What is the best way to approximate continuous functions using sines and cosines ?
  - What is the best way to approximate continuous functions using polynomials ?
  - If we plot the price of houses vs their area (in a particular locality) what is the "best" estimate of price per square foot ?
- These questions fall under the purview of the approximation problem:

The approximation problem

- Consider the following questions:
  - What is the best way to approximate continuous functions using sines and cosines ?
  - What is the best way to approximate continuous functions using polynomials ?
  - If we plot the price of houses vs their area (in a particular locality) what is the "best" estimate of price per square foot ?
- These questions fall under the purview of the approximation problem: Let V be an inner product space and $S \subseteq V$ be a f.d. subspace.

The approximation problem

- Consider the following questions:
  - What is the best way to approximate continuous functions using sines and cosines ?
  - What is the best way to approximate continuous functions using polynomials ?
  - If we plot the price of houses vs their area (in a particular locality) what is the "best" estimate of price per square foot ?
- These questions fall under the purview of the approximation problem: Let V be an inner product space and $S \subseteq V$ be a f.d. subspace. Given an element $x \in V$, determine an element $s \in S$ whose distance from x is as small as possible.

The approximation theorem

The approximation theorem

- Let $S \subseteq V$ be a f.d. subspace of an inner product space $(V, \langle , \rangle)$ and let $x \in V$.

The approximation theorem

- Let $S \subseteq V$ be a f.d. subspace of an inner product space $(V, \langle , \rangle)$ and let $x \in V$. Then if s is the projection of x on S,

The approximation theorem

- Let $S \subseteq V$ be a f.d. subspace of an inner product space $(V, \langle , \rangle)$ and let $x \in V$. Then if s is the projection of x on S, $\|x - s\| \leq \|x - t\|$ for any $t \in S$.

The approximation theorem

- Let $S \subseteq V$ be a f.d. subspace of an inner product space $(V, \langle , \rangle)$ and let $x \in V$. Then if s is the projection of x on S, $\|x - s\| \leq \|x - t\|$ for any $t \in S$. Equality holds if and only if $t = s$.

The approximation theorem

- Let $S \subseteq V$ be a f.d. subspace of an inner product space $(V, \langle , \rangle)$ and let $x \in V$. Then if s is the projection of x on S, $\|x - s\| \leq \|x - t\|$ for any $t \in S$. Equality holds if and only if $t = s$.
- Proof: Note that $x = s + s^{\perp}$ and hence $x - t = (s - t) + s^{\perp}$.

The approximation theorem

- Let $S \subseteq V$ be a f.d. subspace of an inner product space $(V, \langle , \rangle)$ and let $x \in V$. Then if s is the projection of x on S, $\|x - s\| \leq \|x - t\|$ for any $t \in S$. Equality holds if and only if $t = s$.
- Proof: Note that $x = s + s^{\perp}$ and hence $x - t = (s - t) + s^{\perp}$. So $\|x - t\|^2 = \|s - t\|^2 + \|x - s\|^2 \geq \|x - s\|^2$ with equality holding if and only if $s = t$.

Examples

Examples

- Let $V = C[0, 2\pi]$ and S be the space spanned by $\phi_0 = \frac{1}{\sqrt{2\pi}}, \phi_1 = \frac{\cos(x)}{\sqrt{\pi}}, \phi_2 = \frac{\sin(x)}{\sqrt{\pi}}, \ldots, \phi_{2n}$.

Examples

- Let $V = C[0, 2\pi]$ and S be the space spanned by $\phi_0 = \frac{1}{\sqrt{2\pi}}, \phi_1 = \frac{\cos(x)}{\sqrt{\pi}}, \phi_2 = \frac{\sin(x)}{\sqrt{\pi}}, \ldots, \phi_{2n}$.
- The best approximation of $f \in V$ by S is given by the projection $f_n = \sum_k \langle f, \phi_k \rangle \phi_k$ where $\langle f, \phi_k \rangle = \int_0^{2\pi} f \phi_k dx$.

Examples

- Let $V = C[0, 2\pi]$ and S be the space spanned by $\phi_0 = \frac{1}{\sqrt{2\pi}}, \phi_1 = \frac{\cos(x)}{\sqrt{\pi}}, \phi_2 = \frac{\sin(x)}{\sqrt{\pi}}, \ldots, \phi_{2n}$.
- The best approximation of $f \in V$ by S is given by the projection $f_n = \sum_k \langle f, \phi_k \rangle \phi_k$ where $\langle f, \phi_k \rangle = \int_0^{2\pi} f \phi_k dx$. These numbers are called the Fourier coefficients of f.

Examples

- Let $V = C[0, 2\pi]$ and S be the space spanned by $\phi_0 = \frac{1}{\sqrt{2\pi}}, \phi_1 = \frac{\cos(x)}{\sqrt{\pi}}, \phi_2 = \frac{\sin(x)}{\sqrt{\pi}}, \ldots, \phi_{2n}$.
- The best approximation of $f \in V$ by S is given by the projection $f_n = \sum_k \langle f, \phi_k \rangle \phi_k$ where $\langle f, \phi_k \rangle = \int_0^{2\pi} f \phi_k dx$. These numbers are called the Fourier coefficients of f.
- Let $V = C[-1, 1]$ and S be the space spanned by $1, x, \ldots, x^n$.

Examples

- Let $V = C[0, 2\pi]$ and S be the space spanned by $\phi_0 = \frac{1}{\sqrt{2\pi}}, \phi_1 = \frac{\cos(x)}{\sqrt{\pi}}, \phi_2 = \frac{\sin(x)}{\sqrt{\pi}}, \ldots, \phi_{2n}$.
- The best approximation of $f \in V$ by S is given by the projection $f_n = \sum_k \langle f, \phi_k \rangle \phi_k$ where $\langle f, \phi_k \rangle = \int_0^{2\pi} f \phi_k dx$. These numbers are called the Fourier coefficients of f.
- Let $V = C[-1, 1]$ and S be the space spanned by $1, x, \ldots, x^n$. The normalised Legendre polynomials $\psi_0 = \frac{1}{\sqrt{2}}, \psi_1 = \frac{\sqrt{3}}{\sqrt{2}} x, \psi_2 = \frac{\sqrt{5}}{2\sqrt{2}}(3x^2 - 1), \ldots, \psi_n$ form an orthonormal basis for S.

Examples

- Let $V = C[0, 2\pi]$ and S be the space spanned by $\phi_0 = \frac{1}{\sqrt{2\pi}}, \phi_1 = \frac{\cos(x)}{\sqrt{\pi}}, \phi_2 = \frac{\sin(x)}{\sqrt{\pi}}, \ldots, \phi_{2n}$.
- The best approximation of $f \in V$ by S is given by the projection $f_n = \sum_k \langle f, \phi_k \rangle \phi_k$ where $\langle f, \phi_k \rangle = \int_0^{2\pi} f \phi_k dx$. These numbers are called the Fourier coefficients of f.
- Let $V = C[-1, 1]$ and S be the space spanned by $1, x, \ldots, x^n$. The normalised Legendre polynomials $\psi_0 = \frac{1}{\sqrt{2}}, \psi_1 = \frac{\sqrt{3}}{\sqrt{2}} x, \psi_2 = \frac{\sqrt{5}}{2\sqrt{2}}(3x^2 - 1), \ldots, \psi_n$ form an orthonormal basis for S.
- The best polynomial approximation of $f \in V$ by S is given by $\tilde{f}_n = \sum_k \langle f, \psi_k \rangle \psi_k$.

Examples

- Let $V = C[0, 2\pi]$ and S be the space spanned by $\phi_0 = \frac{1}{\sqrt{2\pi}}, \phi_1 = \frac{\cos(x)}{\sqrt{\pi}}, \phi_2 = \frac{\sin(x)}{\sqrt{\pi}}, \ldots, \phi_{2n}$.
- The best approximation of $f \in V$ by S is given by the projection $f_n = \sum_k \langle f, \phi_k \rangle \phi_k$ where $\langle f, \phi_k \rangle = \int_0^{2\pi} f \phi_k dx$. These numbers are called the Fourier coefficients of f.
- Let $V = C[-1, 1]$ and S be the space spanned by $1, x, \ldots, x^n$. The normalised Legendre polynomials $\psi_0 = \frac{1}{\sqrt{2}}, \psi_1 = \frac{\sqrt{3}}{\sqrt{2}}x, \psi_2 = \frac{\sqrt{5}}{2\sqrt{2}}(3x^2 - 1), \ldots, \psi_n$ form an orthonormal basis for S.
- The best polynomial approximation of $f \in V$ by S is given by $\tilde{f}_n = \sum_k \langle f, \psi_k \rangle \psi_k$. For instance, if $f(x) = \sin(\pi x)$, then $\langle f, \psi_0 \rangle = 0$, $\langle f, \psi_1 \rangle = \frac{2\sqrt{3}}{\pi\sqrt{2}}$, $\langle f, \psi_2 \rangle = 0$.

Least squares fit

Least squares fit

- Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$.

Least squares fit

- Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$.
- Unfortunately, in real life, if one collects data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, they will not all lie on a line!

Least squares fit

- Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$.
- Unfortunately, in real life, if one collects data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, they will not all lie on a line!
- We want to find those m, c such that the corresponding line is the "best fit", i.e., $\sum_i (y_i - mx_i - c)^2$ is the smallest possible.

Least squares fit

- Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$.
- Unfortunately, in real life, if one collects data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, they will not all lie on a line!
- We want to find those m, c such that the corresponding line is the "best fit", i.e., $\sum_i (y_i - mx_i - c)^2$ is the smallest possible.
- Note that if we consider the subspace in $\mathbb{R}^n$ spanned by the vectors $(x_1, \ldots, x_n)$ and $(1, 1, \ldots, 1)$,

Least squares fit

- Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$.
- Unfortunately, in real life, if one collects data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, they will not all lie on a line!
- We want to find those m, c such that the corresponding line is the "best fit", i.e., $\sum_i (y_i - mx_i - c)^2$ is the smallest possible.
- Note that if we consider the subspace in $\mathbb{R}^n$ spanned by the vectors $(x_1, \ldots, x_n)$ and $(1, 1, \ldots, 1)$, we essentially want the vector s lying in this space that is the best approximation of the vector $(y_1, y_2, \ldots, y_n)$.

Least squares fit

- Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$.
- Unfortunately, in real life, if one collects data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, they will not all lie on a line!
- We want to find those m, c such that the corresponding line is the "best fit", i.e., $\sum_i (y_i - mx_i - c)^2$ is the smallest possible.
- Note that if we consider the subspace in $\mathbb{R}^n$ spanned by the vectors $(x_1, \ldots, x_n)$ and $(1, 1, \ldots, 1)$, we essentially want the vector s lying in this space that is the best approximation of the vector $(y_1, y_2, \ldots, y_n)$. We can calculate m, c using the formulae directly, or

Least squares fit

- Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$.
- Unfortunately, in real life, if one collects data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, they will not all lie on a line!
- We want to find those m, c such that the corresponding line is the "best fit", i.e., $\sum_i (y_i - mx_i - c)^2$ is the smallest possible.
- Note that if we consider the subspace in $\mathbb{R}^n$ spanned by the vectors $(x_1, \ldots, x_n)$ and $(1, 1, \ldots, 1)$, we essentially want the vector s lying in this space that is the best approximation of the vector $(y_1, y_2, \ldots, y_n)$. We can calculate m, c using the formulae directly, or using the equations $(Y - X\beta)^T X = 0$ where Y is the column vector of y_i,

Least squares fit

- Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$.
- Unfortunately, in real life, if one collects data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, they will not all lie on a line!
- We want to find those m, c such that the corresponding line is the "best fit", i.e., $\sum_i (y_i - mx_i - c)^2$ is the smallest possible.
- Note that if we consider the subspace in $\mathbb{R}^n$ spanned by the vectors $(x_1, \ldots, x_n)$ and $(1, 1, \ldots, 1)$, we essentially want the vector s lying in this space that is the best approximation of the vector $(y_1, y_2, \ldots, y_n)$. We can calculate m, c using the formulae directly, or using the equations $(Y - X\beta)^T X = 0$ where Y is the column vector of y_i, X is the $n \times 2$ matrix whose first column is x and the second column is $(1, 1, \ldots)$, and $\beta = \begin{bmatrix} m \\ 1 \end{bmatrix}$.

Least squares fit

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)
- More generally, if $y = m_1 x_1 + m_2 x_2 + \ldots + m_k x_k + c$, then the "data matrix" X is

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)
- More generally, if $y = m_1x_1 + m_2x_2 + \ldots + m_kx_k + c$, then the "data matrix" X is an $n \times (k+1)$ matrix, and the "slopes vector" β is a $(k+1)$-vector.

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)
- More generally, if $y = m_1x_1 + m_2x_2 + \ldots + m_kx_k + c$, then the "data matrix" X is an $n \times (k+1)$ matrix, and the "slopes vector" β is a $(k+1)$-vector.
- Even then, the principle is to project y onto the "column space", i.e.,

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)
- More generally, if $y = m_1x_1 + m_2x_2 + \ldots + m_kx_k + c$, then the "data matrix" X is an $n \times (k+1)$ matrix, and the "slopes vector" β is a $(k+1)$-vector.
- Even then, the principle is to project y onto the "column space", i.e., the subspace of $\mathbb{R}^n$ generated by the columns of X.

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)
- More generally, if $y = m_1x_1 + m_2x_2 + \ldots + m_kx_k + c$, then the "data matrix" X is an $n \times (k+1)$ matrix, and the "slopes vector" β is a $(k+1)$-vector.
- Even then, the principle is to project y onto the "column space", i.e., the subspace of $\mathbb{R}^n$ generated by the columns of X. Alternatively, $(Y - X\beta)^T X = 0$.

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)
- More generally, if $y = m_1x_1 + m_2x_2 + \ldots + m_kx_k + c$, then the "data matrix" X is an $n \times (k+1)$ matrix, and the "slopes vector" β is a $(k+1)$-vector.
- Even then, the principle is to project y onto the "column space", i.e., the subspace of $\mathbb{R}^n$ generated by the columns of X. Alternatively, $(Y - X\beta)^T X = 0$. These equations are called *normal equations*.

Least squares fit

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)
- More generally, if $y = m_1x_1 + m_2x_2 + \ldots + m_kx_k + c$, then the "data matrix" X is an $n \times (k+1)$ matrix, and the "slopes vector" β is a $(k+1)$-vector.
- Even then, the principle is to project y onto the "column space", i.e., the subspace of $\mathbb{R}^n$ generated by the columns of X. Alternatively, $(Y - X\beta)^T X = 0$. These equations are called *normal equations*. This procedure is called *linear regression*.

- It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.)
- More generally, if $y = m_1x_1 + m_2x_2 + \ldots + m_kx_k + c$, then the "data matrix" X is an $n \times (k+1)$ matrix, and the "slopes vector" β is a $(k+1)$-vector.
- Even then, the principle is to project y onto the "column space", i.e., the subspace of $\mathbb{R}^n$ generated by the columns of X. Alternatively, $(Y - X\beta)^T X = 0$. These equations are called *normal equations*. This procedure is called *linear regression*.
- By the way, if you want to fit polynomials, you can do exactly the same thing by the trick of introducing new variables ! $(x_1 = x, x_2 = x^2, \ldots)$.