

1. Projections

(a) False Consider $\mathbf{b}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, they form a basis of \mathbf{R}^2 . When using the definition $\mathcal{P}_S x = \sum_{i=1}^n \langle x, b_i \rangle b_i$ we would expect that $\mathcal{P}_S b_1 = b_1$. However $\mathcal{P}_S b_1 = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \neq b_1$.

(b) True Let $S^\perp = \{x | \langle x, y \rangle = 0, \forall y \in S\}$ a subspace of an inner product space X , then $S^{\perp\perp} = \{x | \langle x, y \rangle = 0, \forall y \in S^\perp\}$. The inner product being symmetric, $S \subseteq S^{\perp\perp}$. Since for any vector $x \in X$, we have $x = y + z$ where $y \in S, z \in S^\perp$, using Gram-schmidt orthonormalization process, we can find a basis of S and S^\perp which express any vector of X as a linear combination of these two basis and combining these two basis together forms a new basis for X so $\dim X = \dim S + \dim S^\perp$. If $\dim X = n$ and $\dim S = m$ then $\dim S^\perp = n - m$. Similarly $\dim S^{\perp\perp} = n - (n - m) = m$ so $\dim S^{\perp\perp} = \dim S$, so $S^{\perp\perp} \subseteq S$ and since the dimension of a space or subspace is the cardinality of its basis, thus $S = S^{\perp\perp}$.

(c) True consider $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$, we want $\mathbf{w} = \begin{bmatrix} \frac{\sum_{i=1,n} v_i}{n} \\ \vdots \\ \frac{\sum_{i=1,n} v_i}{n} \end{bmatrix}$. The orthogonal

projection of \mathbf{v} onto the vector \mathbf{b} is defined as $\frac{\mathbf{v} \cdot \mathbf{b}}{\|\mathbf{b}\|^2}$, take $\mathbf{b} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.

2. Eigen decomposition Rewriting the problem in a matrix form:

$$\begin{pmatrix} d_{n+1} \\ w_{n+1} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 5 & -3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} d_n \\ w_n \end{pmatrix}$$

Let $A = \frac{1}{4} \begin{pmatrix} 5 & -3 \\ 1 & 1 \end{pmatrix}$, $v_{n+1} = \begin{pmatrix} d_{n+1} \\ w_{n+1} \end{pmatrix}$, $v_0 = \begin{pmatrix} d_0 \\ w_0 \end{pmatrix}$ then $v_{n+1} = A v_n = A^n v_0$. We are looking to find the eigen decomposition so we can understand the behavior of v_n as $n \rightarrow \infty$. $\det(A - \lambda I) = \frac{1}{2}(2\lambda^2 - 3\lambda + 1)$, we find for eigenvalues $\lambda_1 = \frac{1}{2}$ and $\lambda_2 = 1$ with corresponding eigenvectors

$w_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $w_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$. Since A is diagonalizable the vectors $\{w_1, w_2\}$ forms a basis of \mathbf{R}^2 and we can express v_0 in this basis as $v_0 = \alpha w_1 + \beta w_2$ for some $\alpha, \beta \in \mathbf{R}$, thus $v_{n+1} = \alpha A^n w_1 + \beta A^n w_2 = \alpha \lambda_1^n w_1 + \beta \lambda_2^n w_2 = \alpha (\frac{1}{2})^n w_1 + \beta w_2$. Then taking the $n \rightarrow \infty$, the first term goes to zero and $v_{n+1} \sim \beta w_2$. So asymptotically $\frac{d_{n+1}}{w_{n+1}} \sim 3$ which verifies the initial condition: $w_0 < d_0$.

3. Function approximation

(a) Using Gram-Schmidt orthonormalization process, we find

$$\begin{aligned}
 v_1 &= 1 \\
 v_2 &= x - \langle x, 1 \rangle \frac{1}{\langle 1, 1 \rangle} \\
 &= x \\
 v_3 &= x^2 - \langle x^2, v_2 \rangle \frac{v_2}{\langle v_2, v_2 \rangle} - \langle x^2, v_1 \rangle \frac{v_1}{\langle v_1, v_1 \rangle} \\
 &= x^2 - \frac{1}{3}
 \end{aligned}$$

Then we normalize each of these vectors to obtain:

$$\begin{aligned}
 w_1 &= \frac{v_1}{\|v_1\|} = \frac{\sqrt{2}}{2} \\
 w_2 &= \frac{v_2}{\|v_2\|} = \sqrt{\frac{3}{2}} x \\
 w_3 &= \frac{v_3}{\|v_3\|} = \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3}\right)
 \end{aligned}$$

(b) The projection of $f(x) = \cos(\frac{\pi}{2} x)$ in the orthonormal basis $\{w_1, w_2, w_3\}$ is: $\sum_{i=1,3} \langle f, w_i \rangle w_i$, where:

$$\begin{aligned}
 \langle f, w_1 \rangle &= \int_{-1}^1 \cos\left(\frac{\pi}{2} x\right) \frac{\sqrt{2}}{2} dx \\
 &= \frac{4}{\pi\sqrt{2}} \sim 0.9 \\
 \langle f, w_2 \rangle &= \int_{-1}^1 \cos\left(\frac{\pi}{2} x\right) \frac{\sqrt{3}}{2} x dx \\
 &= 0 \\
 \langle f, w_3 \rangle &= \int_{-1}^1 \cos\left(\frac{\pi}{2} x\right) \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3}\right) dx \\
 &= 2\sqrt{10} \frac{\pi^2 - 12}{\pi^3} \sim -0.43
 \end{aligned}$$

4. Scalar linear approximation

(a) First we write $E[(ax + b - y)^2] = E[((ax - y) - (-b))^2]$, we know that the best mean-squared error minimizer of a random variable is its mean

so $-b = E[ax - y] = a E[x] - E[y] = a\mu_x - \mu_y$. Substituting b in the expression we want to minimize gives us:

$$\begin{aligned} E[(ax + b - y)^2] &= E[(ax - y - (a\mu_x - \mu_y))^2] \\ &= E[\{a(\mu_x - x) - (y - \mu_y)\}^2] \\ &= a^2 E[(x - \mu_x)^2] + E[(y - \mu_y)^2] - 2a E[(x - \mu_x)(y - \mu_y)] \\ &= a^2 \sigma_x^2 + \sigma_y^2 - 2a \text{Cov}(x, y) \end{aligned}$$

Let $f(a) = a^2 \sigma_x^2 + \sigma_y^2 - 2a \text{Cov}(x, y)$, then $f'(a) = 2(\sigma_x^2 a - \text{Cov}(x, y))$ and $f''(a) = 2\sigma_x^2$. The function is strictly convex, and its second derivative is positive, thus its minimizer is $a = \frac{\text{Cov}(x, y)}{\sigma_x^2} = \rho_{x, y} \frac{\sigma_y}{\sigma_x}$.

5. Gradients

- (a) Compute the gradient of $f(x) = b^T x$ where $b \in \mathbf{R}^d$ and $f : \mathbf{R}^d \rightarrow \mathbf{R}$.
 $\frac{\partial f(x)}{\partial x_j} = \sum_i b_i \frac{\partial x_i}{\partial x_j} = b_j$, thus $\nabla f(x) = b$.
- (b) Compute the gradient of $f(x) = x^T A x$ where $A \in \mathbf{R}^{d \times d}$ and $f : \mathbf{R}^d \rightarrow \mathbf{R}$. $f(x) = x^T A x = \sum_{i=1}^d \sum_{j=1}^d a_{ij} x_i x_j$, then

$$\begin{aligned} \frac{\partial f}{\partial x_k} &= \sum_{i=1}^d \sum_{j=1}^d a_{ij} \frac{\partial x_i x_j}{\partial x_k} \\ &= \sum_{i=1}^d \sum_{j=1}^d a_{ij} (x_j \delta_{ik} + x_i \delta_{jk}) \\ &= \sum_{i=1}^d \sum_{j=1}^d a_{ij} x_j \delta_{ik} + \sum_{i=1}^d \sum_{j=1}^d a_{ij} x_i \delta_{jk} \\ &= \sum_{j=1}^d a_{kj} x_j + \sum_{i=1}^d a_{ik} x_i \\ &= (Ax)_k + (Ax)_k^T \end{aligned}$$

thus $\nabla f(x) = (A + A^T)x$.