Homework 0

Solutions

1. (Projections)

(a) False. This only holds if the vectors in the basis are orthogonal. Take

$$b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{1}$$

This is obviously a basis of $\mathbb{R}^2$. However

$$\begin{aligned} \mathcal{P}_{\mathbb{R}^2}\, b_1 = b_1 \neq \sum_{i=1}^{m} \langle b_1, b_i \rangle\, b_i \\ = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ = \begin{bmatrix} 1 \\ 2 \end{bmatrix}. \end{aligned}$$

(b) True. We will prove that each set includes the other set, which implies that they are equal.

$\mathcal{S} \subseteq \left(\mathcal{S}^{\perp}\right)^{\perp}$. If $x \in \mathcal{S}$ then for any $v \in \mathcal{S}^{\perp}$ $x \perp v$ by the definition of orthogonal complement, so $x \in \left(\mathcal{S}^{\perp}\right)^{\perp}$.

$\left(\mathcal{S}^{\perp}\right)^{\perp} \subseteq \mathcal{S}$. Let $x \in \left(\mathcal{S}^{\perp}\right)^{\perp}$. Let us assume $\mathcal{S}$ has dimension d. Let $b_1, \ldots, b_d$ be an orthonormal basis of $\mathcal{S}$, and $b_{d+1}, \ldots, b_n$ a basis of $\mathcal{S}^{\perp}$. The set $b_1, \ldots, b_n$ is an orthonormal basis for all of $\mathbb{R}^n$ (for $i \leq d$ and $j > d$ $\langle b_i, b_j \rangle = 0$ because $b_i \in \mathcal{S}$ and $b_i \in \mathcal{S}^{\perp}$), so

$$x = \sum_{i=1}^{n} \langle b_i, x \rangle\, b_i. \tag{2}$$

However x is orthogonal to $\mathcal{S}^{\perp}$ so $\langle b_i, x \rangle = 0$ for $d+1 \leq i \leq n$, which implies that x is in the span of $b_1, \ldots, b_d$ and consequently belongs to $\mathcal{S}$.

(c) For any $x \in \mathbb{R}^n$,

$$\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x[i] \\ \cdots \\ \frac{1}{n}\sum_{i=1}^{n} x[i] \end{bmatrix} = \left\langle \frac{1}{\sqrt{n}}\vec{1}, x \right\rangle \frac{1}{\sqrt{n}}\vec{1}, \tag{3}$$

where $\vec{1}$ is the vector of ones. Since $\|\vec{1}\|_2^2 = n$, $\frac{1}{\sqrt{n}}\vec{1}$ is a unit norm vector, so this is a projection onto its span.

2. (Eigendecomposition) Using the eigendecomposition of the matrix corresponding to the system of equations, we obtain that the population of deer and wolfs at time n is given by

$$\begin{aligned}
\begin{bmatrix} d_n \\ w_n \end{bmatrix} &= \frac{1}{4^n} \begin{bmatrix} 5 & -3 \\ 1 & 1 \end{bmatrix}^n \begin{bmatrix} d_0 \\ w_0 \end{bmatrix} && (4) \\
&= \frac{1}{4^n} \left(\begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \right)^n \begin{bmatrix} d_0 \\ w_0 \end{bmatrix} && (5) \\
&= \frac{1}{4^n} \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}^n \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} d_0 \\ w_0 \end{bmatrix} && (6) \\
&= \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{4^n} \end{bmatrix} \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 1.5 \end{bmatrix} \begin{bmatrix} d_0 \\ w_0 \end{bmatrix}. && (7)
\end{aligned}$$

Taking the limit as $n \to \infty$ yields

$$\lim_{n\to\infty} \begin{bmatrix} d_n \\ w_n \end{bmatrix} = \frac{d_0 - w_0}{2} \begin{bmatrix} 3 \\ 1 \end{bmatrix}. \tag{8}$$

There will be three times more deer than wolfs, as long as $d_0 > w_0$.

3. (Function approximation)

(a) Let $u_1, u_2, u_3 \in P_2$ denote the orthonormal basis produced by Gram-Schmidt. Then

$$\begin{aligned}
u_1 &= \frac{1}{\|1\|} = \frac{1}{\sqrt{2}} \approx 0.7071 && (9) \\
\langle x, u_1 \rangle &= \frac{1}{\sqrt{2}} \int_{-1}^{1} x \, dx = 0 && (10) \\
x - \langle x, u_1 \rangle u_1 &= x && (11) \\
u_2 &= \frac{x}{\|x\|} = \sqrt{\frac{3}{2}} x \approx 1.22474x && (12) \\
\langle x^2, u_1 \rangle &= \frac{1}{\sqrt{2}} \int_{-1}^{1} x^2 \, dx = \frac{\sqrt{2}}{3} && (13) \\
\langle x^2, u_2 \rangle &= \sqrt{\frac{3}{2}} \int_{-1}^{1} x^3 \, dx = 0 && (14) \\
x^2 - \langle x^2, u_1 \rangle u_1 - \langle x^2, u_2 \rangle u_2 &= x^2 - 1/3 && (15) \\
u_3 &= \frac{x^2 - 1/3}{\|x^2 - 1/3\|} = \sqrt{\frac{45}{8}} (x^2 - 1/3) && (16) \\
&\approx -0.7906 + 2.3717x^2. && (17)
\end{aligned}$$

(b) We have

$$\langle f, u_1 \rangle = \frac{2\sqrt{2}}{\pi} \approx 0.9003 \tag{18}$$

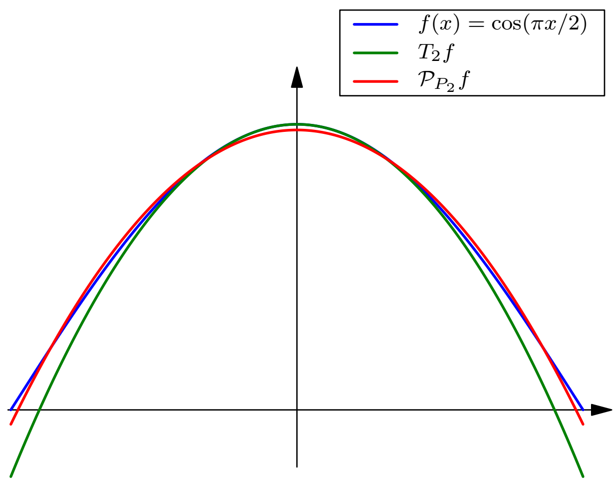
$$\langle f, u_2 \rangle = 0 \tag{19}$$

$$\langle f, u_3 \rangle = \frac{2\sqrt{10}\,(\pi^2 - 12)}{\pi^3} \approx -0.4346 \tag{20}$$

$$\mathcal{P}_{P_2} f = \sum_{i=1}^{3} \langle u_i, f \rangle u_i \tag{21}$$

$$= \frac{15\,(\pi^2 - 12)\left(x^2 - \frac{1}{3}\right)}{\pi^3} + \frac{2}{\pi} \approx 0.9802 - 1.0306x^2. \tag{22}$$

(c) We have $T_2 f(x) = 1 - \pi^2 x^2/8$.



(d) The projection is the best approximator in the norm induced by the inner product, so in particular $\|f - \mathcal{P}_{P_2} f\| < \|f - T_2 f\|$. The Taylor expansion provides a local approximation, which is arbitrarily close to the function as we approach 0, but can deviate significantly in other regions.

4. (Scalar linear estimation)

(a) By Lemma 2.1, if we fix a the optimal b equals $E\left(\tilde{y} - a\tilde{x}\right) = \mu_{\tilde{y}} - a\mu_{\tilde{x}}$. Plugging this in we have

$$\min_{b \in \mathbb{R}} \mathrm{E}[(a\tilde{x} + b - \tilde{y})^2] = \mathrm{E}[(a(\tilde{x} - \mu_{\tilde{x}}) - (\tilde{y} - \mu_y))^2] \tag{23}$$

$$= a^2 \mathrm{Var}(\tilde{x}) + \mathrm{Var}(\tilde{y}) - 2a\mathrm{Cov}(\tilde{x}\tilde{y}). \tag{24}$$

This is a convex quadratic function with respect to a, so we can set the derivative to zero to find the minimum:

$$a_{\text{opt}} = \frac{\mathrm{Cov}(\tilde{x}\tilde{y})}{\mathrm{Var}(\tilde{x})} \tag{25}$$

$$= \frac{\sigma_{\tilde{y}} \rho_{\tilde{x},\tilde{y}}}{\sigma_{\tilde{x}}}, \tag{26}$$

so

$$b_{\text{opt}} = \mu_{\tilde{y}} - \frac{\sigma_{\tilde{y}} \rho_{\tilde{x},\tilde{y}} \mu_{\tilde{x}}}{\sigma_{\tilde{x}}}. \tag{27}$$

(b) By independence,

$$\begin{aligned} \mathrm{E}(\tilde{x}\tilde{y}) &= \mathrm{E}(\tilde{z})\mathrm{E}(\tilde{y}^2) && (28) \\ &= 0. && (29) \end{aligned}$$

The estimate equals

$$a_{\text{opt}}\tilde{x} + b_{\text{opt}} = \mu_{\tilde{y}}. \tag{30}$$

(c) Set $\tilde{z} = 1$ with probability 1/2, and $\tilde{z} = -1$ with probability 1/2. Then if $\tilde{x} > 0$, we set the estimate to $\tilde{x}$, and if $\tilde{x} < 0$, we set the estimate to $-\tilde{x}$.

5. (Gradients)

(a)

$$\begin{aligned} \frac{\partial f(x)}{\partial x[i]} &= \frac{\partial \sum_{j=1}^{n} b[j]x[j]}{\partial x[i]} && (31) \\ &= b[i], && (32) \end{aligned}$$

so the gradient equals b.

(b)

$$\begin{aligned} \frac{\partial f(x)}{\partial x[i]} &= \frac{\partial \sum_{j=1}^{n} \sum_{k=1}^{n} A[j,k]x[j]x[k]}{\partial x[i]} && (33) \\ &= \sum_{j=1}^{n} A[i,j]x[j] + \sum_{j=1}^{n} A[j,i]x[j], && (34) \end{aligned}$$

so the gradient equals $Ax + A^T x$.