Homework 9 Solutions

1. (15 points) *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 $x \in \left(\mathcal{S}^{\perp}\right)^{\perp}$. By Lemma in Lecture notes 9 we can write

$$\mathcal{P}_{\mathcal{S}}\, x = x - \mathcal{P}_{\mathcal{S}^{\perp}}\, x, \tag{2}$$

but

$$\begin{aligned} \left|\left|\mathcal{P}_{\mathcal{S}^{\perp}}\, x\right|\right|_{\langle\cdot,\cdot\rangle} &= \langle x - \mathcal{P}_{\mathcal{S}}\, x, \mathcal{P}_{\mathcal{S}^{\perp}}\, x \rangle & (3) \\ &= 0, & (4) \end{aligned}$$

because x is orthogonal to $\mathcal{S}^{\perp}$ and so is $\mathcal{P}_{\mathcal{S}}\, x$. As a result, $\mathcal{P}_{\mathcal{S}^{\perp}}\, x = 0$ and therefore $x = \mathcal{P}_{\mathcal{S}}\, x$. This proves that $x \in \mathcal{S}$.

c. False. If we initialize using v_2 then the power method will just stay in the span of v_2.

2. (10 points) *Heart beat*

a. Note that

$$\begin{aligned} \mathrm{E}\left(Y_1^2\right) &= (1+3)\sigma^2 = 4\sigma^2, & (5) \\ \mathrm{E}\left(Y_2^2\right) &= (3+1)\sigma^2 = 4\sigma^2, & (6) \\ \mathrm{E}\left(Y_1 Y_2\right) &= 3\sigma^2 = 3\sigma^2, & (7) \\ \mathrm{E}\left(B Y_1\right) &= \sigma^2, & (8) \\ \mathrm{E}\left(B Y_2\right) &= 0 & (9) \end{aligned}$$

We apply Gram-Schmidt to obtain an orthonormal basis consisting of the random variables U_1 and U_2:

$$U_1 = \frac{Y_1}{||Y_1||} \tag{10}$$
$$= \frac{Y_1}{2\sigma}, \tag{11}$$
$$V_1 = Y_2 - \left\langle Y_2, \frac{Y_1}{||Y_1||} \right\rangle \frac{Y_1}{||Y_1||} \tag{12}$$
$$= Y_2 - \frac{\mathrm{E}\left(Y_1 Y_2\right)}{4\sigma^2} Y_1 \tag{13}$$
$$= Y_2 - \frac{3Y_1}{4} \tag{14}$$
$$U_2 = \frac{V_2}{||V_2||} \tag{15}$$
$$= \frac{2Y_2 - 1.5Y_1}{\sqrt{7}\sigma}. \tag{16}$$

where we have used

$$\left|\left| Y_2 - \frac{3Y_1}{4} \right|\right|^2 = \mathrm{E}\left(\left(Y_2 - \frac{3Y_1}{4}\right)\left(Y_2 - \frac{3Y_1}{4}\right)\right) \tag{17}$$
$$= \mathrm{E}\left(Y_2^2\right) + \frac{9\,\mathrm{E}\left(Y_1^2\right)}{16} - \frac{3\,\mathrm{E}\left(Y_1 Y_2\right)}{2} \tag{18}$$
$$= 4\sigma^2 + \frac{9\sigma^2}{4} - \frac{9\sigma^2}{2} \tag{19}$$
$$= \frac{7\sigma^2}{4} \tag{20}$$

b. The projection is

$$\mathcal{P}_{\mathrm{span}(\{Y_1,Y_2\})}\, B = \langle U_1, B\rangle\, U_1 + \langle U_2, B\rangle\, U_2 \tag{21}$$
$$= \frac{\mathrm{E}\left(BY_1\right) Y_1}{4\sigma^2} - \frac{1.5\,\mathrm{E}\left(BY_1\right)\left(2Y_2 - 1.5Y_1\right)}{7\sigma^2} \tag{22}$$
$$= \frac{Y_1}{4} - \frac{12Y_2 - 9Y_1}{28} \tag{23}$$
$$= \frac{4Y_1}{7} - \frac{3Y_2}{7}. \tag{24}$$

The linear estimate cancels out M to some extent while limiting the amount of noise. This achieves better MSE than just using Y_1 or $Y_1 - Y_2$.

3. (10 points) *Cold (10 points).*

a. To alleviate notation let H_k denote the event that Bob is healthy on day k and I_k the event that he is ill. The model described in the problem can be written as a Markov chain,

$$\begin{bmatrix} \mathrm{P}\left(H_k\right) \\ \mathrm{P}\left(I_k\right) \end{bmatrix} = \begin{bmatrix} \mathrm{P}\left(\text{healthy}|\text{healthy}\right) & \mathrm{P}\left(\text{healthy}|\text{ill}\right) \\ \mathrm{P}\left(\text{ill}|\text{healthy}\right) & \mathrm{P}\left(\text{ill}|\text{ill}\right) \end{bmatrix} \begin{bmatrix} \mathrm{P}\left(H_{k-1}\right) \\ \mathrm{P}\left(I_{k-1}\right) \end{bmatrix}$$
$$= \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} \mathrm{P}\left(H_{k-1}\right) \\ \mathrm{P}\left(I_{k-1}\right) \end{bmatrix}. \tag{25}$$

If we assume that the model has been valid for a long time, then the Markov chain will have converged to its stationary distribution, the eigenvector corresponding to the eigenvalue 1 divided by its sum:

$$\pi_\infty = \frac{v_1}{\sum_i^n v_1[i]}. \tag{26}$$

$$\begin{bmatrix} \mathrm{P}\left(\text{healthy}\right) \\ \mathrm{P}\left(\text{ill }\right) \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}. \tag{27}$$

The probability that Bob has a cold is consequently 1/3.

b. Using the Markov model we compute

$$\begin{aligned}
\mathrm{P}\left(I_k|I_{k-3}\right) &= \mathrm{P}\left(I_k, H_{k-1}, H_{k-2}|I_{k-3}\right) + \mathrm{P}\left(I_k, H_{k-1}, I_{k-2}|I_{k-3}\right) && (28)\\
&\quad + \mathrm{P}\left(I_k, I_{k-1}, H_{k-2}|I_{k-3}\right) + \mathrm{P}\left(I_k, I_{k-1}, I_{k-2}|I_{k-3}\right) && (29)\\
&= \mathrm{P}\left(I_k|H_{k-1}\right)\mathrm{P}\left(H_{k-1}|H_{k-2}\right)\mathrm{P}\left(H_{k-2}|I_{k-3}\right) && (30)\\
&\quad + \mathrm{P}\left(I_k|H_{k-1}\right)\mathrm{P}\left(H_{k-1}|I_{k-2}\right)\mathrm{P}\left(I_{k-2}|I_{k-3}\right) && (31)\\
&\quad + \mathrm{P}\left(I_k|I_{k-1}\right)\mathrm{P}\left(I_{k-1}|H_{k-2}\right)\mathrm{P}\left(H_{k-2}|I_{k-3}\right) && (32)\\
&\quad + \mathrm{P}\left(I_k|I_{k-1}\right)\mathrm{P}\left(I_{k-1}|I_{k-2}\right)\mathrm{P}\left(I_{k-2}|I_{k-3}\right) && (33)\\
&= 0.1 \cdot 0.9 \cdot 0.2 + 0.1 \cdot 0.2 \cdot 0.8 + 0.8 \cdot 0.1 \cdot 0.2 + 0.8 \cdot 0.8 \cdot 0.8 && (34)\\
&= 0.562. && (35)
\end{aligned}$$

Now, to compute the MAP estimator

$$\begin{aligned}
\mathrm{P}\left(I_{k-3}|I_k\right) &= \frac{\mathrm{P}\left(I_k|I_{k-3}\right)\mathrm{P}\left(I_{k-3}\right)}{\mathrm{P}\left(I_k\right)} && (36)\\
&= \mathrm{P}\left(I_k|I_{k-3}\right) \qquad \text{if we have reached the stationary distribution by } k-3 && (37)\\
&= 0.562. && (38)
\end{aligned}$$

By the Law of Total Probability $\mathrm{P}\left(H_{k-3}|I_k\right) = 0.438$. The MAP estimate is that Bob was ill three days ago.

4. (10 points) *Wheat (15 points).*

a. The code is

```
U, s, V = np.linalg.svd(data, full_matrices=True)
projections = np.dot(np.transpose(U[:,:2]),data)
projections_1 = projections[:,ind_x_eq_val(labels,1)]
projections_2 = projections[:,ind_x_eq_val(labels,2)]
projections_3 = projections[:,ind_x_eq_val(labels,3)]
```

The projections are shown in Figure 1. When we visualize the projection on the two first two principal components the different classes are separated into clusters. This is not the case when we project onto the two last principal components.

b. There are many possible answers to this question. An option is to perform PCA separately on the three classes and select the first k principal components of each for a certain k. Then we can project the new data vector onto the principal components of each class and select the class for which the projection has higher norm.

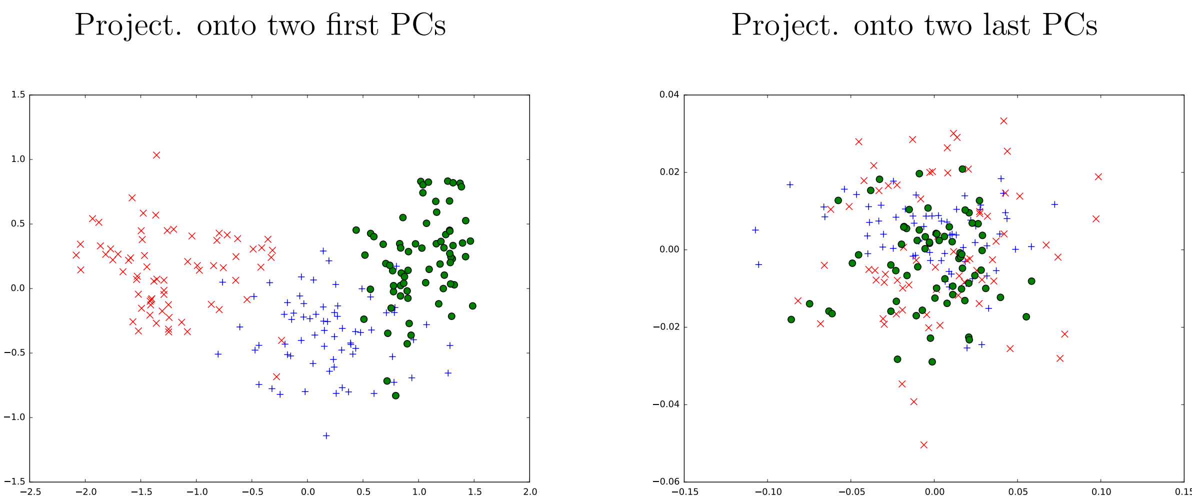

Figure 1: Projection of the wheat data onto the two first and the two last principal components.