CSE 250B: Homework 2 Solutions

1. *Error rate of 1-NN classifier.*

   (a) Consider a training set in which the same point x appears twice, but with different labels. The training error of 1-NN on this data will not be zero.

   (b) We mentioned in class that the risk of the 1-NN classifier, $R(h_n)$, approaches $2R^*(1-R^*)$ as $n \to \infty$ where R^* is the Bayes risk. If $R^* = 0$, this means that the 1-NN classifier is consistent: $R(h_n) \to 0$.

2. *Bayes optimality in a multi-class setting.* The Bayes-optimal classifier predicts the label that is most likely:

$$h^*(x) = \arg\max_{i \in |\mathcal{Y}|} \eta_i(x)$$

3. *Classification with an abstain option.* The classifier should abstain whenever the probability of error exceeds θ:

$$h^*(x) = \begin{cases} \text{abstain} & \text{if } \theta < \eta(x) < 1-\theta \\ 1 & \text{if } \eta(x) \geq 1-\theta \\ 0 & \text{if } \eta(x) \leq \theta \end{cases}$$

4. *The statistical learning assumption.*

   (a) Here, μ is the distribution over proposed songs, while η tells us which songs will be successful. Both are likely to change with time, violating the statistical learning assumption. However, the drift might be quite slow, so a classifier trained today may work well for another year or two before needing to be re-trained.

   (b) In this example, the bank's data set consists only of loans it *accepted*. It is not a random sample from μ, which is the distribution over all loan applications. This is a severe violation of the i.i.d. sampling requirement.

   (c) The move from the west coast to the entire country means that μ is changing, and it is possible that η is changing as well. Technically, this violates the statistical learning assumption; but it is possible that the change in distribution may not be very severe.

5. *Conditional probability.*

   (a) He is most likely to be in `happy` mood.

   (b) The probability of the baby being happy is $\Pr(\texttt{happy}|\text{talks a little})$.

$$\begin{aligned}\Pr(\texttt{happy}|\text{talks a little}) &= \frac{\Pr(\text{talks a little}|\texttt{happy})\Pr(\texttt{happy})}{\Pr(\text{talks a little})} \\ &= \frac{\Pr(\text{talks a little}|\texttt{happy})\Pr(\texttt{happy})}{\Pr(\text{talks a little}|\texttt{happy})\Pr(\texttt{happy}) + \Pr(\text{talks a little}|\texttt{sad})\Pr(\texttt{sad})} \\ &= \frac{\frac{1}{6} \times \frac{3}{4}}{\frac{1}{6} \times \frac{3}{4} + \frac{1}{6} \times \frac{1}{4}} \;=\; \frac{3}{4}\end{aligned}$$

   Therefore, the probability of the prediction begin wrong is $1 - \Pr(\texttt{happy}|\text{talks a little}) = 1/4$.

6. *Bayes optimal classifier.*

   (a)

$$h^*(x) = \arg\min_{i \in \mathcal{Y}} \pi_i P_i(x) = \begin{cases} 1 & \text{if } -1 \leq x \leq 0 \\ 3 & \text{if } 0 < x \leq 1 \end{cases}$$

(b) The probability density function of $\mathcal{X}$ is

$$\mu(x) = \begin{cases} \frac{13}{24} & x \in [-1, 0] \\ \frac{11}{24} & x \in (0, 1] \end{cases}$$

Looking at all the ways to be wrong, the error rate is

$$\Pr(y = 1 \text{ and } x > 0) + \Pr(y = 2) + \Pr(y = 3 \text{ and } x \leq 0) \;=\; \frac{1}{3} \cdot \frac{1}{8} + \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{2} \;=\; \frac{11}{24}$$

7. $(1/\sqrt{14}, 2/\sqrt{14}, 3/\sqrt{14})$
8. $(-1/\sqrt{2}, 1/\sqrt{2})$ and $(1/\sqrt{2}, -1/\sqrt{2})$
9. $x \cdot x = 25 \Leftrightarrow \|x\| = 5$. All points of length 5: a sphere, centered at the origin, of radius 5.
10. $f(x) = 2x_1 - x_2 + 6x_3 = w \cdot x$ for $w = (2, -1, 6)$.
11. A is 10×30 and B is 30×20
12. (a) X is $n \times d$
    (b) XX^T is $n \times n$
    (c) $(XX^T)_{ij} = x^{(i)} \cdot x^{(j)}$
13. $((x^T x)(x^T x)(x^T x)) = (\|x\|^2)^3 = 10^6$
14. $x^T x = \|x\|^2 = 35$ and

$$xx^T = \begin{bmatrix} 1 & 3 & 5 \\ 3 & 9 & 15 \\ 5 & 15 & 25 \end{bmatrix}$$

15. The angle θ between x and y satisfies $\cos\theta = x^T y / \|x\|\|y\| = 1/2$, so θ is 60 degrees.
16.

$$M = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 0 & 0 \\ -2 & 0 & 6 \end{bmatrix}$$

17. *Symmetric Matrices*
    (a) $(AA^T)^T = (A^T)^T A^T = AA^T$, Thus AA^T is symmetric.
    (b) $(A^T A)^T = A^T (A^T)^T = A^T A$, Thus $A^T A$ is symmetric.
    (c) $(A + A^T)^T = (A^T + A) = (A + A^T)$, Thus $(A + A^T)$ is symmetric
    (d) $(A - A^T)^T = (A^T - A) \neq (A - A^T)$, Thus $(A - A^T)$ need not be symmetric
18. (a) $|A| = 8! = 40320$
    (b) $A^{-1} = \text{diag}(1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8)$
19. *Orthonormal matrices*
    (a) UU^T is the identity matrix
    (b) $U^{-1} = U^T$
20. Since A is singular matrix, $|A| = 0 \implies z - 6 = 0 \implies z = 6$