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. *Covariance and correlation.*

   (a) The covariance matrix is $\begin{pmatrix} 100 & 200 \\ 200 & 400 \end{pmatrix}$

   (b) $\text{corr}(X, Y) = \frac{\text{cov}(X,Y)}{\text{std}(X)\text{std}(Y)} = 1$

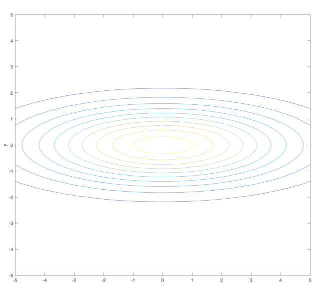
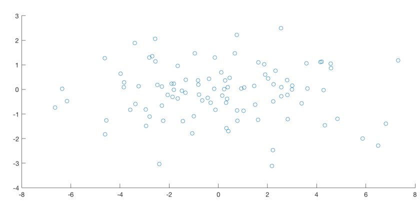
8. *Bivariate Gaussians.*

   (a) $\mu = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 0.25 \end{pmatrix}$.

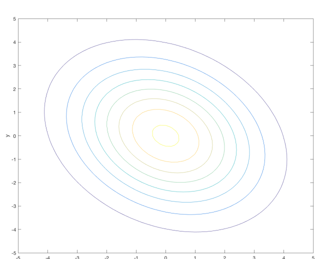
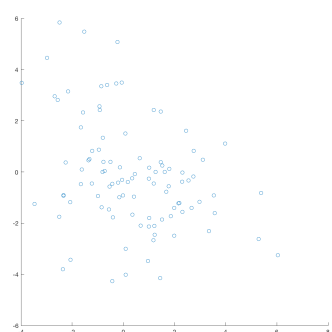
   (b) $\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

9. *More bivariate Gaussians.*

   (a) Contour lines and scatter plot:

   (b) Contour lines and scatter plot:

10. *Qualitative appraisal of Gaussian parameters.*

   (a) If the two variables are negatively correlated, then $b < 0$.

   (b) If the two variables are uncorrelated, then $b = 0$.

   (c) If one variable is the linear function of another, then $b = \pm\sqrt{ac}$.

   (d) If one of the variables is a constant, then $b = 0$ and $ac = 0$.