

Note: The problems in this assignment are to be solved by hand. If you want to use a computer to help with the plots, you may; but you will probably find that unnecessary.

1. In a 2-class problem with 1 feature, you are given the following data points:

$$S_1 : -3, -2, 0$$

$$S_2 : -1, 2$$

- (a) Give the k -nearest neighbor estimate of $p(x|S_1)$ with $k=3$, for all x . Give both algebraic (simplest) form, and a plot. On the plot, show the location (x value) and height of (each) peak.
- (b) Give the Parzen Windows estimate of $p(x|S_2)$ with window function:

$$\Delta(u) = \begin{cases} 0.25, & -2 \leq u < 2 \\ 0, & \text{otherwise} \end{cases}$$

Give both algebraic (simplest) form, and a plot. On the plot, show the x value of all significant points, and label the values of $p(x|S_2)$ clearly.

- (c) Estimate prior probabilities $P(S_1)$ and $P(S_2)$ from frequency of occurrence of the data points.
- (d) Give an expression for the decision rule for a Bayes minimum error classifier using the density and probability estimates from (a)-(c). You may leave your answer in terms of $\hat{p}(x|S_1)$, $\hat{p}(x|S_2)$, $\hat{P}(S_1)$, $\hat{P}(S_2)$, without plugging in for these quantities.
- (e) Using the estimates you have made above, solve for the decision boundaries and regions of a Bayes minimum error classifier using the density and probability estimates from (a)-(c). Give your answer in 2 forms:
- Algebraic expressions of the decision rule, in simplest form (using numbers and variable x);
 - A plot showing the decision boundaries and regions.
- Tip:** you may find it easiest to develop the algebraic solution and plot at the same time.
- (f) Classify the points: $x = -0.5, 0.1, 0.5$ using the classifier you developed in (e).
- (g) Separately, use a discriminative 3-NN classifier to classify the points $x = -0.5, 0.1, 0.5$. (**Hint:** if this takes you more than a few steps for each data point, you are doing more work than necessary.)

Assignment continues on next page...

2. [Comment: this problem is on parameter estimation, which is covered in Lecture 26 on Monday, 4/27.]

In a 1D problem (1 feature), we will estimate parameters for one class. We model the density $p(x|\theta)$ as:

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

in which $\theta \geq 0$.

You are given a dataset $Z: x_1, x_2, \dots, x_N$, which are drawn i.i.d. from $p(x|\theta)$.

In this problem, you may use for convenience the notation:

$$m \triangleq \frac{1}{N} \sum_{i=1}^N x_i.$$

- (a) Solve for the maximum likelihood (ML) estimate $\hat{\theta}_{ML}$, of θ , in terms of the given data points. Express your result in simplest form.

For parts (b) and (c) below, assume there is a prior for θ , as follows:

$$p(\theta) = \begin{cases} a e^{-a\theta}, & \theta \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

in which $a \geq 0$.

- (b) Solve for the maximum a posteriori (MAP) estimate $\hat{\theta}_{MAP}$, of θ , in terms of the given data points. Express your result in simplest form.
- (c) Write $\hat{\theta}_{MAP}$ as a function of $\hat{\theta}_{ML}$ and given parameters. Find $\lim_{\sigma_\theta \rightarrow \infty} \hat{\theta}_{MAP}$, in which σ_θ is the standard deviation of the prior on θ . What does this limit correspond to in terms of our prior knowledge of θ ?

Hint: the standard deviation of θ for the given $p(\theta)$ is: $\sigma_\theta = \frac{1}{a}$.

3. **[Extra credit]** Comment: this problem is not more difficult than the regular-credit problems above; it is extra credit because the total length Problems 1 and 2 above is already sufficient and reasonable for one homework assignment.

In a 2-class problem with D features, you are to use Fisher's Linear Discriminant to find an optimal 1D feature space. You are given that the scatter matrices for each class (calculated from the data for each class) are diagonal:

$$\underline{S}_1 = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_D^2 \end{bmatrix}, \quad \underline{S}_2 = \begin{bmatrix} \rho_1^2 & & & 0 \\ & \rho_2^2 & & \\ & & \ddots & \\ 0 & & & \rho_D^2 \end{bmatrix}$$

and you are given the sample means for each class:

$$\underline{m}_1 = \begin{pmatrix} m_1^{(1)} \\ m_2^{(1)} \\ \vdots \\ m_D^{(1)} \end{pmatrix}, \quad \underline{m}_2 = \begin{pmatrix} m_1^{(2)} \\ m_2^{(2)} \\ \vdots \\ m_D^{(2)} \end{pmatrix}.$$

- Find the Fisher's Linear Discriminant \underline{w} . Express in simplest form.
- Let $D = 2$. Suppose $\sigma_1^2 = 4\sigma_2^2$, and $\rho_1^2 = 4\rho_2^2$, and:

$$\underline{m}_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \underline{m}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

Plot vectors \underline{m}_1 , \underline{m}_2 , $(\underline{m}_1 - \underline{m}_2)$, and \underline{w} .

- Interpreting your answer of part (b), which makes more sense for a 1D feature space direction: $(\underline{m}_1 - \underline{m}_2)$ or \underline{w} ? Justify your answer.