Solutions to Optional Homework (Lecture 15)

Problem 12.1

(a) By inspection, the mean vectors of the three classes are, approximately, $\mathbf{m}_1 = (1.5, 0.3)^T$, $\mathbf{m}_2 = (4.3, 1.3)^T$, and $\mathbf{m}_3 = (5.5, 2.1)^T$ for the classes Iris setosa, versicolor, and virginica, respectively. The decision functions are of the form given in Eq. (12-4). Substituting the preceding values of mean vectors gives:

$$d_1(\mathbf{x}) = \mathbf{x}^T\mathbf{m}_1 - \frac{1}{2}\mathbf{m}_1^T\mathbf{m}_1 = 1.5x_1 + 0.3x_2 - 1.2$$
$$d_2(\mathbf{x}) = \mathbf{x}^T\mathbf{m}_2 - \frac{1}{2}\mathbf{m}_2^T\mathbf{m}_2 = 4.3x_1 + 1.3x_2 - 10.1$$
$$d_3(\mathbf{x}) = \mathbf{x}^T\mathbf{m}_3 - \frac{1}{2}\mathbf{m}_3^T\mathbf{m}_3 = 5.5x_1 + 2.1x_2 - 17.3$$

(b) The decision boundaries are given by the equations

$$d_{12}(\mathbf{x}) = d_1(x) - d_2(x) = -2.8x_1 - 1.0x_2 + 8.9 = 0$$
$$d_{13}(\mathbf{x}) = d_1(x) - d_3(x) = -4.0x_1 - 1.8x_2 + 16.1 = 0$$
$$d_{23}(\mathbf{x}) = d_2(x) - d_3(x) = -1.2x_1 - 0.8x_2 + 7.2 = 0$$

Figure P12.1 shows a plot of these boundaries.

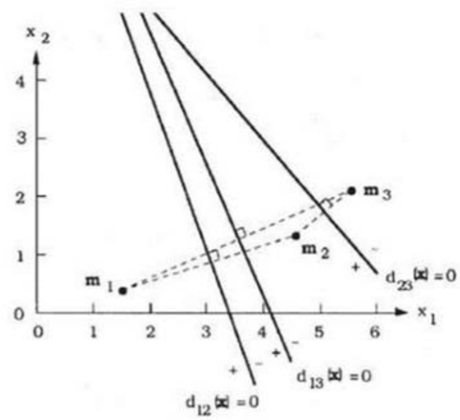

Figure P12.1

Problem 12.9

(a) Because it is given that the pattern classes are governed by Gaussian densities, only knowledge of the mean vector and covariance matrix of each class are needed to specify the Bayes classifier. Substituting the given patterns into Eqs. (12-29) and (12-30) we obtain

$$\mathbf{m}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mathbf{m}_2 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{C}_1^{-1}$$

and

$$\mathbf{C}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{C}_2^{-1} = \mathbf{C}_1^{-1}$$

Because $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{I}$, the decision functions are the same as those of a minimum-distance classifier:

$$d_1(\mathbf{x}) = \mathbf{x}^T \mathbf{m}_1 - \frac{1}{2}\mathbf{m}_1^T \mathbf{m}_1 = 1.0x_1 + 1.0x_2 - 1.0$$

$$d_2(\mathbf{x}) = \mathbf{x}^T \mathbf{m}_2 - \frac{1}{2}\mathbf{m}_2^T \mathbf{m}_2 = 5.0x_1 + 5.0x_2 - 25$$

The decision boundary is given by the equation $d(\mathbf{x}) = d_1(\mathbf{x}) - d_2(\mathbf{x}) = 0$, or

$$d(\mathbf{x}) = -4x_1 - 4x_2 + 24 = 0$$

(b) Figure P12.9 shows this boundary in pattern space.

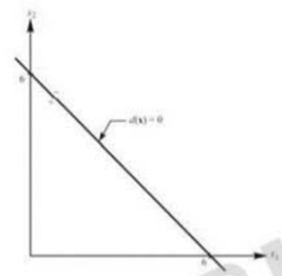

Figure p12.9

Problem 12.17

The single decision function that implements a minimum-distance classifier for two classes in n-dimensional space has the form

$$d_{12}(\mathbf{x}) = d_1(\mathbf{x}) - d_2(\mathbf{x}) = (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{x} - \frac{1}{2}\left(\mathbf{m}_1{}^T \mathbf{m}_1 - \mathbf{m}_2{}^T \mathbf{m}_2\right)$$

A single neuron with n inputs performs the computation

$$\begin{aligned} z &= \sum_{k=1}^{n} w_k a_k + b \\ &= \boldsymbol{w}^T \mathbf{a} + b \end{aligned}$$

As you can see, this is exactly the same form as above, so out neural network is a two-layer network: an input layer whose values are the value of the input vector $\mathbf{x}$, (so that $a_k = x_k$) and an output layer consisting of a single neuron with weights

$$w_k = m_{1k} - m_{2k}; \quad k = 1, 2, \ldots, n$$

and bias

$$b = -\frac{1}{2}\left(\mathbf{m}_1{}^T \mathbf{m}_1 - \mathbf{m}_2{}^T \mathbf{m}_2\right)$$

The network has no hidden layers. If we use a tanh activation function, after the weights and bias are learned via training, this neural network will have an output >0 for patterns of one class and <0 for patterns of the other class. This is exactly how a minimum distance classifier with a single decision function would behave for two pattern classes that are linearly separable. Because they are tightly grouped, we assume each mean vector is a good representation of one of the classes.

Problem 12.23

Figure P12.23 shows the solution.

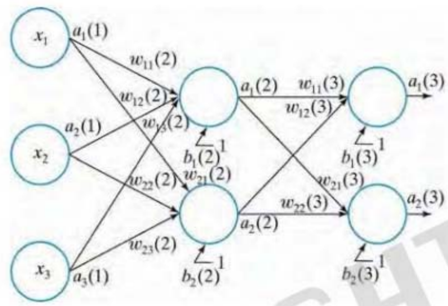

Figure P12.23

Problem 12.30

(a) If the kernels are square, $w \times w$, of odd size, and the square dimensions of the convolution planes is 504, then $(512 - 2 \cdot (w-1)/2) = 504$, so the spatial dimensions of the kernel are 9×9.

(b) Because the spatial dimensions of the subsampling kernels are 2×2, the size of the feature planes is reduced by 50%. Therefore, the pooled feature planes are of size $(504 * 0.5) \times (504 * 0.5) = 252 \times 252$.

(c) The depth (number) of pooled feature maps in a layer is equal to the number of feature maps in that layer: 12.

(d) $(252-2) \times (252-2) = 250 \times 250$.

(e) We know from (d) that the dimensions of the feature maps in the second layer are 250×250 elements. The pooling neighborhoods are of size 2×2, so the dimensions of the pooled feature maps are 125×125. The number of pooled feature maps is the same as the number of feature maps, which we are given is 6. When vectorized, each pooled feature map will yield of vector of dimension $(125)^2 \times 1 = 15{,}625 \times 1$. There are 6 such maps, so that when all 6 vectors are "stacked" we will end up a vector of dimension $93{,}750 \times 1$ being input into the fully-connected neural net.