EE 660 **Homework Week 1** Posted: Wed., 8/22/2018
Jenkins Due: Tues., 8/28/2018, 2:00 PM*

## Please note

1. *Turn in your assignment by uploading a single pdf file to the assignment dropbox, on D2L (Content Area, Week 1). Please keep your file size reasonable (less than 2 MB).
2. These are reading problems; that is, they are based on the reading, which is done (for the most part) *before* the material is covered in lecture. Also, the answers to reading problems are usually short.

## Reading

*Introduction:* Murphy Ch. 1 (lightly is OK).

*Regression, part 1:* Murphy 1.4.5 (Linear Regression introduction), and 7.1 - 7.5, inclusive. The sections with asterisks (for example 7.4*) are optional; 7.3.2 is also optional (even though it has no asterisk). You might prefer to omit the optional sections on the first reading; then afterwards read any optional sections you are interested in.

## Problems

Note that single underbar (as in $\underline{x}$) denotes a vector quantity, and double underbar (as in $\underline{\underline{X}}$ ) denotes a matrix quantity (here and throughout our class).

1. (Murphy Ch. 1 and lecture) Suppose for the iris classification problem, we want to classify just 2 types: Virginica ($y = +1$) and Versicolor ($y = -1$), and we use the same 4 features (petal length, petal width, sepal length, sepal width). Our dataset $\mathcal{D}$ consists of values from $N$ flowers, thus:

$$\mathcal{D} = \{\underline{x}_i, y_i\}_{i=1}^{N}$$

   (a) Write the design matrix $\underline{\underline{X}}$ and output vector $\underline{y}$ in terms of $\underline{x}_i$ and $y_i$.

   (b) What is the dimension of $\underline{\underline{X}}$ ?

   (c) Suppose instead we use only two features (petal length, petal width). Let scalar $x_1$ denote petal length, and scalar $x_2$ denote petal width. Suppose we use a linear classifier; write an expression for the function $\hat{y}(\underline{x})$, in terms of $x_1, x_2, w_0, w_1, w_2$. Would a linear classifier be able to classify most of the dataset correctly? (Refer to Fig. 1.4 for the dataset.) **No need** to run this on computer or to find values for the weights.

**For Problems 2 and 3 below**, consider a stock-price regression problem, in which we are given the stock price at each day's close of the stock exchange over the past year; this comprises our dataset $\mathcal{D} = \left\{\left(x_i, y_i\right)\right\}_{i=1}^{N}$. We want to use regression to estimate the values of the stock at other times of days in the past year, and to predict the value at various times and days in the future.

2. (Murphy Sec. 7.2 and 1.4.5) Let's model the estimated function as a $d$-degree polynomial, of a scalar variable $x$:

$$\hat{f}\left(x\right) = \sum_{k=0}^{d} w_k x^k \tag{1}$$

in which $d$ is given. $x$ is time. (For example, $x$ could be measured in units of days, allowing for fractional days, so that $x \in \mathbb{R}$ (real numbers). Our training data is then limited to $1 \le x_i \le 365$, with $x_i \in \mathbb{Z}$ (integers).)

(a) Give an expression for each component of $\underline{\phi}\left(x\right)$, the basis function expansion of $x$. What is the dimension of $\underline{\phi}\left(x\right)$?

(b) What variables in Eq. (1) above are to be found using a learning algorithm (or optimization procedure)?

3. (Murphy Sec. 7.3) Continuing the model of Problem 2,

(a) In Eq. (7.4), what is the parameter vector to be estimated, $\underline{\theta}$, in the notation of our stock-price problem?
**Hint:** if not sure, then continue reading and answer later.

(b) What assumption does Murphy make on $p\left(y \middle| x, \underline{\theta}\right)$, and on the training samples?

4. (Murphy Sec. 7.5) Fill in the steps to get from Eq. (7.31) to Eq. (7.32) in Murphy.
**Hint:** the Normal or Gaussian density function is given in Murphy 2.4.1.