Homework From Week 9: Linear Nets, Stats

1. **Breast Cancer Data Classifier**

   Think about what the numbers mean in terms of the classes we've described, and give a short summary.

   SOLUTION: See the attached "published" script files. Both methods (finally) gave similar answers, but we definitely had to re-scale our input for the Widrow-Hoff update rule (otherwise, we got a lot of numerical overflows). Once re-scaled, a learning rate of about 0.01 and a number of epochs of about 500 gave fairly good results.

   The confusion matrix is interesting in this example, as it shows that "Carcinoma" was classified correctly. Unfortunately, we also see that some other types were *also* classified as carcinoma (not a good thing to happen, this is a "false positive" or Type I error). Class 3 is especially problematic, since it is not classified correctly a majority of the time.

   Another interesting feature is that Classes 5 and 6 are sometimes confused for each other, but are never confused for Classes 1-4. Therefore, these form their own "cluster" away from the rest.

   From the Statistics chapter, Exercises 10 and 11, written again below.

10. Show that, for fixed constants a, b, $\text{Cov}(\mathbf{x} + a, \mathbf{y} + b) = \text{Cov}(\mathbf{x}, \mathbf{y})$

    SOLUTION: First, if $\bar{x}$, $\bar{y}$ are the means of data in vectors $\mathbf{x}$, $\mathbf{y}$ respectively, then the means of $\mathbf{x} + a$ and $\mathbf{y} + b$ are simply $\bar{x} + a$ and $\bar{y} + b$. Therefore:

$$\text{Cov}(\mathbf{x} + a, \mathbf{y} + b) = \frac{1}{p-1} \sum_{k=1}^{p} (\mathbf{x} + a) - (\bar{x} + a))((\mathbf{y} + b) - (\bar{y} + b))$$

$$= \frac{1}{p-1} \sum_{k=1}^{p} (\mathbf{x} - \bar{x})(\mathbf{y} - \bar{y}) = \text{Cov}(\mathbf{x}, \mathbf{y})$$

11. As a summary of question 11, if $\mathbf{y} = m\mathbf{x} + b$, then find the covariance and correlation coefficient between $\mathbf{x}$ and $\mathbf{y}$.

    SOLUTION:

$$\text{Cov}(\mathbf{x}, m\mathbf{x} + b) = \frac{1}{p-1} \sum_{k=1}^{p} (\mathbf{x} - \bar{x})((m\mathbf{x} + b) - (m\bar{x} + b)) =$$

$$\frac{m}{p-1} \sum_{k=1}^{p} (\mathbf{x} - \bar{x})^2 = m\text{Var}(\mathbf{x}) = m s_x^2$$

For the correlation coefficient,

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{m s_x^2}{s_x s_y}$$

Now the variance of $\mathbf{y}$ is given by:

$$s_y^2 = \frac{1}{p-1}\sum_{k=1}^{p}((m\mathbf{x}+b)-(m\bar{x}+b))^2 = \frac{1}{p-1}\sum_{k=1}^{p} m^2(\mathbf{x}-\bar{x})^2 = m^2 s_x^2$$

Therefore,

$$s_y = \sqrt{m^2} s_x = |m| s_x$$

Substituting this back in, we get:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{m s_x^2}{s_x |m| s_x} = \frac{m}{|m|} = \begin{cases} 1 & \text{if } m > 0 \\ -1 & \text{if } m < 0 \end{cases}$$

(From the Linear Regression Section, last page)

2. Find the line of best fit through the data found in the table on p. 39 of the linear regression notes (this is the Hanford data relating an "index" to the number of deaths). Show the result by graphing the data and the line you found.

   Since $A^T A$ is invertible, you can solve the equation given, $A\mathbf{c} = \mathbf{t}$ using Matlab. The plotting commands are also given below, assuming the slope is the first coordinate of $\mathbf{c}$ and the intercept is the second coordinate, and the "index" is the first column of A.

```
x=...   %Enter the data as a column vector
t=...   %Enter the data as a column vector
A=...   %Construct the matrix A
c=inv(A'*A)*A'*t;
xx=linspace(min(x),max(x));
yy=c(1)*xx+c(2);
plot(x,t,'r*',xx,yy,'k-');
```

3. Be sure you understand how to apply the Widrow-Hoff update rule. For example, suppose W is the 2×2 identity matrix, and the other vectors are given by the following, and the learning rate $\alpha = \frac{1}{10}$.

$$\mathbf{b} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \mathbf{t} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

   Perform one step of the Widrow-Hoff rule to update W and $\mathbf{b}$.

   SOLUTION: We need $\mathbf{y} = W\mathbf{x} + \mathbf{b}$ to do the update. That's easy enough to do in our heads: $\mathbf{y} = [-1, 2]^T$. Now the update rule is to compute:

$$W + \alpha(\mathbf{t}-\mathbf{y})\mathbf{x}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{10}\begin{bmatrix} 3 \\ -3 \end{bmatrix}[0,1] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{10}\begin{bmatrix} 0 & 3 \\ 0 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 3/10 \\ 0 & 7/10 \end{bmatrix}$$

$$\mathbf{b} + \alpha(\mathbf{t}-\mathbf{y}) = (7/10)[-1, 1]^T$$