Intelligent Data Analysis

Problem Set 5 **Due 5/12/2016**

1. Derive the ridge regression estimates given in Chapter 8, pp 18. Every input variable x_i has zero mean. The error function is given by

$$E_\lambda\left(\left\{\mathbf{w}_j\right\}_{j=0}^{d}\right)=\sum_{t=1}^{N}\left(r^t-w_0-\sum_{j=1}^{d}w_jx_j^t\right)^2+\lambda\sum_{j=1}^{d}w_j^2 .$$

   (a) Let $\mathbf{X}$ be the $N\times d$ data matrix. Show that the ridge regression estimates minimizing the error function are $w_0=\frac{1}{N}\sum_{t=1}^{N}r^t$ and $\mathbf{w}=\left[w_1\cdots w_d\right]^T=\left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I}_d\right)^{-1}\mathbf{X}^T\mathbf{r}$.

   (b) If the error function is modified as

$$E'_\lambda\left(\left\{\mathbf{w}_j\right\}_{j=0}^{d}\right)=\sum_{t=1}^{N}\left(r^t-w_0-\sum_{j=1}^{d}w_jx_j^t\right)^2+\lambda\sum_{j=1}^{d}c_jw_j^2 ,$$

   where c_j are positive constants. Find w_0 and $\mathbf{w}$ that minimizes $E'_\lambda\left(\left\{\mathbf{w}_j\right\}_{j=0}^{d}\right)$.

2. See Chapter 16, pp 23. Prove the following statements:
   (a) $H(Y\,|\,X)=H(X,Y)-H(X)$
   (b) $I(X;Y)=H(X)+H(Y)-H(X,Y)=H(Y)-H(Y\,|\,X)$

3. Program Assignment
   Download the image of R. A. Fisher from
   http://upload.wikimedia.org/wikipedia/commons/4/46/R._A._Fischer.jpg.
   (a) Show the original image (268 x 326 pixels). Perform image compression using 2x2 block vector quantization, as shown in Chapter 9, pp 21. Show the compressed images using 4 and 32 code vectors, respectively, and determine the compression rate.
   (b) Use SVD to approximate the image. Let the image be denoted by $\mathbf{X}$. Perform SVD and obtain $\mathbf{X}=\mathbf{U\Sigma V}^T$. Let $\mathbf{X}_k=\mathbf{U\Sigma}_k\mathbf{V}^T$, where $\mathbf{\Sigma}_k$ keeps the k largest singular values of $\mathbf{\Sigma}$ and sets all the remaining singular values to zero. Show the approximated images for k=4 and 10.

4. <u>Program Assignment</u>

Download the Prostate data from e3. See Chapter 6, pp 41 for detailed description of the data set. You are asked to do data survey. Use the whole data set (containing 97 cases) to answer the following questions. For the output variable (lpas) and each of the input variables, divides the state space into *B*=7 equally sized regions. Compute the entropy for each variable in the data set and determine its norm, and then the whole set entropy, as shown in Chapter 16, pp 24. Compute the conditional entropy $H(Y \mid X_i)$ and the mutual information $I(X_i;Y)$, for $i = 1,\ldots,8$. Also show their norms. See Chapter 16, pp 25-26.