Machine Learning - Assignment 7

Posted Saturday November 24, 2007
Due Monday december 3, 2007

1. [20 points] **The theoretical details of PCA**

   In class we discussed in detail the case of PCA for finding one component, and we summarized what happens if k components are needed. Now, we analyze this in more detail. Suppose that the linear projection onto a k-dimensional subspace that maximizes the variance accounted for is defined by the k largest eigenvalues of the scatter matrix. We need to show that this is true for $k+1$ as well (then it will be true by induction). To do this, consider the variance and set the derivative of the variance wrt the new vector v_{k+1} to be 0. This should be done under the constraints that v_{k+1} be orthogonal to $v_1 \dots v_k$, and that it also be normalized to unit length. Use Lagrange multipliers to enforce these constraints. Then, use the orthonormality properties of the vectors to show that v_{k+1} must be an eigenvector of the scatter matrix. Finally, show that variance is maximized if v_{k+1} corresponds to λ_{k+1}.

2. [20 points] **Understanding correlations**

   In this problem, we will show that the independence of two random variables is a sufficient but not necessary condition for the correlation matrix to be diagonal.

   (a) [10 points] Consider two random variables X and Z which are independent, i.e. $p(X, Z) = p(X)p(Z)$. Show that any off-diagonal elements in the correlation matrix must be 0.

   (b) [10 points] Now suppose that $Z = X^2$, and $X\tilde{U}nif[-1. + 1]$. Write down $p(Z|X)$. Show that all off-diagonal elements in the correlation matrix must be 0, by using the fact that $p(X, Z) = p(Z|X)p(X)$

3. [60 points] **Playing with PCA**

   (a) [5 points] Generate 200 examples from a Gaussian with mean (5, 20) and covariance matrix:

   $$\begin{bmatrix} 10 & 1 \\ -1 & 5 \end{bmatrix}$$

   Plot the data you generated.

   (b) [5 points] Make a prediction about what directions the principal components should have, based on the class notes.

   (c) [10 points] Run PCA on this data and describe what happens.

   (d) [5 points] Now subtract the mean from all the data points and run PCA. Describe again what happens. Is there any difference in the principal components found? Explain the result

(e) [5 points] Now subtract the mean from all the data points and divide the result by the standard deviation (so as to normalize the data). Run PCA again and explain what happens to the result.

(f) [5 points] Multiply the second coordinate of every point in the data set by 1000 and run PCA again. What happens to the principle components, and why?

(g) [5 points] Comment on the robustness of PCA wrt the scaling of the data, and on what needs to be done to make sure that good results are obtained.

(h) [25 points] Generate a new data set from the function $y = x^3 - x^2 + x + 1$ with x in the range of -1 to 1. Sample x uniformly randomly, compute y then add normal noise with mean 0 and standard deviation 1 to y. Generate 100 data points in this way. Run PCA on this two-dimensional data and plot the results. Implement kernel PCA on this data, with a kernel of your choice. Run kernel PCA and report the eigenvectors found. Generate another 10 points and project them onto the principal components in both scenarios, then measure the reconstruction error. Briefly explain what you observe.