Machine Learning. Fall 2013. Homework 3.

Due: Thursday 30/May. 2013.

Question 1. Principal Component Analysis.

Suppose the data elements $\{\vec{x}_i\}$ where each $\vec{x}_i$ is an M-dimensional vector. The vectors are of form $\vec{x} = a\delta_k = (0, ..., 0, a, 0, ...)$, where the a is in the k^{th} slot, and k, a are random variables. k is uniformly distributed over $1, .., N$ and $P(a)$ is arbitrary. Calculate the covariance matrix of the data $\{\vec{x}_i\}$. Show that it has one eigenvector of form $(1, ..., 1)$ and that the other eigenvectors all have the same eigenvalue. Discuss whether PCA is a good way to select features for this problem. Hint: The covariance matrix C of the signals $\vec{x}$ is of form $C_{i,j} = \lambda + \mu\delta_{i,j}$ for some λ, μ.

Question 2. Fisher's linear discriminant.

Describes Fisher's linear discriminant. How is it used to discriminate between data from two classes.

Suppose each datapoint $\vec{x}$ in the first class is of form $\vec{x} = (x_1, ..., x_{2M})$ where the x_i are i.i.d. from a Gaussian with zero mean and standard deviation σ. The datapoints in the second class are of form $\vec{x} = (x_1, ..., x_M, \rho + x_{M+1}, ..., \rho + x_{2M})$ where ρ is fixed and the x_i are also generated by a Gaussian with zero mean and standard deviation σ.

What is Fisher's linear discriminant between these two datasets? Does the discriminant change if ρ is a random variable with distribution $P(\rho)$?

Question 3. ISOMAP algorithm.

Describe the ISOMAP algorithm. What are its advantages and disadvantages compared to PCA?

Question 4. Expectation-Maximization.

Do questions 3 and 4 from Chp 7 of Alypaydin's book.

Question 5. Decision Trees.

Describe the Decision Tree algorithm. Consider the task of deciding whether a customer is low-risk $y = 1$ or high-risk $y = -1$ depending on income x_1 and savings x_2. Suppose the set of questions are tests of form $is\ x_1 > T_1$ and $is\ x_2 > T_2$, where T_1 and T_2 are thresholds. The training set has low-risk $y = 1$ points at $x_1, x_2)$ positions: $(2, 3), (3.5, 4), (2.5, 6), (6, 3.5), (7, 8)$ and high-risk $y = -1$ points at $(7, 1.5), (1, 8), (1.5, 1.5), (2, 2), (3, 3)$. Derive the best decision tree for this case, specifying the impurities at the nodes.