ECE M146 Discussion 8 Solution
Introduction to Machine Learning Friday, May 31, 2019
Instructor: Lara Dolecek
TA: Zehui (Alex) Chen, Ruiyi (John) Wu

1. **Minimum-error formulation of PCA**
   Consider a complete orthonormal set of D-dimensional basis vector $\{u_i\}$ that satisfy $u_i^T u_j = \delta_{ij}$. For any D-dimensional data point x_n, we can represent it with a linear combination of the basis vectors using
$$x_n = \sum_{i=1}^{D} (x_n^T u_i) u_i.$$
   We want to approximate each data point x_n by
$$\tilde{x}_n = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{D} b_i u_i,$$
   while minimizing
$$J = \frac{1}{N} \sum_{n=1}^{N} \|x_n - \tilde{x}_n\|^2.$$
   Find the optimal u_i, z_{ni} and b_i.
   Solution: We first consider the minimization with respect to z_{ni} by setting the derivative to be 0 and we get
$$z_{nj} = x_n^T u_j.$$
   Similarly, setting the derivative of J with respect to b_i to zero, we get
$$b_j = \bar{x}^T u_j.$$
   Substitute $z_n i$ and b_i. We get
$$x_n - \tilde{x}_n = \sum_{i=M+1}^{D} [(x_n - \bar{x})^T u_i] u_i.$$
   With this, we see that find that
$$J = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=M+1}^{D} (x_n^T u_i - \bar{x}^T u_i)^2 = \sum_{i=M+1}^{D} u_i^T S u_i.$$
   What remains is the constrained minimization with respect to $u_i, i = M+1, \cdots, N$. We get these base vectors should be the eigenvectors corresponding to the smallest eigenvalues by induction.

2. **Some intuition on bagging** Consider a regression problem in which we are trying to predict the value of a single continuous variable, and suppose we generate M bootstrap data sets and then use each to train a separate copy $y_m(x)$ of a predictive model where $m = 1, ..., M$. The bagging prediction is given by

$$y_{bagging}(x) = \frac{1}{M}\sum_{m=1}^{M} y_m(x).$$

Suppose the true regression function that we are trying to predict is given by $h(x)$, so that the output of each of the models can be written as the true value plus an error in the form

$$y_m(x) = h(x) + \epsilon_m(x).$$

Assume $E[\epsilon_m(x)] = 0$ and $E[\epsilon_m(x)\epsilon_l(x)] = 0, i \neq l$. Find the average sum-of-squares error made by the models acting individually. Also find the expected error of the bagging predictor.

Solution:

$$E_{AV} = \frac{1}{M}\sum_{m=1}^{M} E[\epsilon_m(x)^2],$$

$$E_{bagging} = E\left[\left(\frac{1}{M}\sum_{m=1}^{M}\epsilon_m(x)\right)^2\right] = \frac{1}{M}E_{AV}.$$