Homework 1

Q1. Show that the ridge regression estimate is the mean (and mode) of the posterior distribution, under a Gaussian prior $\beta \sim N(0, \tau^2\mathbf{I})$, and Gaussian sampling model $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$. Find the relationship between the regularization parameter λ in the ridge formula, and the variances τ^2 and σ^2.

Q2. Show that the ridge regression estimates can be obtained by ordinary least squares regression on an augmented data set. We augment the centered matrix $\mathbf{X}$ with p additional rows $\sqrt{\lambda}\mathbf{I}$ and augment $\mathbf{y}$ with p zeroes. By introducing artificial data having response value zero, the fitting procedure is forced to shrink the coefficients toward zero.

Q3. Consider a mixture model density in p-dimensional feature space,

$$g(x) = \sum_{k=1}^{K} \pi_k g_k(x),$$

where $g_k = \mathcal{N}(\mu_k, \sigma^2\mathbf{I})$ and $\pi_k \geq 0$ for all k with $\sum_k \pi_k = 0$. Here $\{\mu_k, \pi_k\}$, $k = 1, ..., K$ and σ^2 are unknown parameters. Suppose we have data $x_1, ..., x_N \sim g(x)$ and we wish to fit the mixture model.

a. Write down the log-likelihood of the data.

b. Derive an EM algorithm for computing the maximum likelihood estimates. Prove that the EM algorithm converges. Hint: use Jensen's inequality and monotone convergence theorem.

c. Show that if σ has a known value in the mixture model and we take $\sigma \to 0$, then in a sense, this EM algorithm coincides with K-means clustering.

Q4. Derive equation (6.8) in [1, p. 195] for multidimensional x.

Q5. Implement logistic regression for *binary* classification for the MNIST dataset. Here binary classification means classifying whether an image is the digit i or not the digit i, where $i = 0, 1, 2, \ldots, 9$.

References

[1] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*. Springer, New York, 2009.