Practice Midterm Exam Questions

October 16, 2016

1. Which of the statements about regularized least squares problems are true?

   Solution. True: B,C. False: A,D.

   A. Using a quadratic regularizer tends to produce sparse solutions.
   B. Imposing a Gaussian prior on w and maximizing the posterior is equivalent to ridge regression.
   C. We can still find a least squares solution even when some features are linearly dependent.
   D. Using a (nonzero) regularizer always increases the out of sample error.

2. You work in fraud detection at a large credit card company. You're about to roll out a new credit card that you hope will appeal to university students. Your current portfolio of card holders mostly consists of professionals in their 40s and 50s who fly over 50,000 miles a year. You've fit a model to your current portfolio of card holders to predict who will be a good credit risk, and it works quite well. Your manager would like to use it to decide which users should be approved for the new card. Do you think this is a good idea? Why or why not?

   Solution. I'd be concerned about using my model to predict performance for the new card. The applicants for the current portfolio of cards differ systematically from the applicants the company expects for the new card. These groups of applicants are *drawn from different distributions.* A model that is very good at discriminating between university students who will or won't pay their bills is very different from one that distinguishes between older professionals: for example, student debt may matter much more to the first group, while mortgages and divorces matter more for the second. These latter phenomena are not present in my training set, and so my model will not be able to predict performance.

   More formally, if I wrote down a feature matrix X for this problem, I'd have some columns (divorce, mortgage size) that are all 0 for my current card holders. So I cannot estimate the model coefficient for that feature.

3. Suppose we want to solve the smoothed least squares problem

$$\text{minimize} \quad \|y - Xw\|^2 + \|Dw\|^2,$$

with variable w, where $D \in \mathbf{R}^{(d-1)\times d}$ is the smoothing matrix we saw in class,

$$D_{ij} = \begin{cases} 1 & j = i \\ -1 & j = i+1 \\ 0 & \text{else} \end{cases}.$$

Derive a set of equations that hold at the solution w and explain why they hold.

Solution. This problem is smooth, unconstrained, and convex, so the minimum of the objective occurs where the gradient with respect to the variable w is 0. Taking the gradient, we see this occurs when

$$\begin{aligned}
0 &= \nabla(\|y - Xw\|^2 + \|Dw\|^2) \\
&= \nabla((y - Xw)^T(y - Xw) + (Dw)^T Dw) \\
&= \nabla(y^T y - 2w^T X^T y + w^T X^T X w + w^T D^T D w) \\
&= -2X^T y + 2X^T X w + 2D^T D w \\
X^T y &= (X^T X + D^T D)w
\end{aligned}$$

If $X^T X + D^T D$ is invertible, then

$$w = (X^T X + D^T D)^{-1} X^T y.$$

However, since D has more columns than rows, $D^T D$ cannot be full rank, so without knowing more about X we cannot decide whether $X^T X + D^T D$ is invertible.