STA414/2104: Practice Problems 3

January, 2018

These practice problems are not for credit. Students may complete independently, in groups, or however they like. Treat these as representative of what might be on the tests.

Questions involving coding may be completed in any language. If you would like help regarding your language of choice from the course team, use R or Python. There will not be any *code*-related questions on the tests, but you will be asked about comuptational algorithms.

1. For each of the following models, state which is a *linear* model according to definition given in lecture 1 (and lecture 2, and lecture 3):
   (a) $y(\mathbf{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$
   (b) $y(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_d x^d$
   (c) $y(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x^{\beta_2} + \beta_3 x^{\beta_4}$
   (d) $y(\mathbf{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 \log x_2$
   (e) $y(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})}$

2. Show that the $L2$-regularized linear regression model

$$\mathbf{t} = \mathbf{X}\mathbf{w} + \boldsymbol{\epsilon}$$

with $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}_n)$ and corresponding loss function

$$L(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^{n} \left(t_i - \mathbf{w}'\mathbf{x}_i\right)^2 + \frac{\lambda}{2} \mathbf{w}'\mathbf{w}$$

has the same solution $\mathbf{w}^*$ as using the posterior mean from a Bayesian linear regression, where the prior distribution of $\mathbf{w}$ is given by

$$\mathbf{w} \sim N(0, \alpha \mathbf{I}_p)$$

where n is the dimension of $\mathbf{t}$ (sample size), p is the dimension of the data, $\mathbf{I}_p$ is the $p \times p$ identity matrix, and $\alpha \in \mathbb{R}$. In general $\alpha \neq \lambda$; can you find a relationship between α, λ, and σ^2?

3. Let

$$y \sim Binom(n, p)$$

where n is known and p is an unknown parameter. Put a prior on p,

$$p \sim Beta(\alpha, \beta)$$

where the Beta distribution has pdf

$$f(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

Find the posterior distribution of $p|y$. Explain why this means that the Beta distribution is the *conjugate prior* for the Binomial.

4. Consider the LBFM code example discussed in class

   (a) Re-run my code, trying different values of the random seed in the cross-validation procedure. Notice how the results are very unstable. Comment on why this might be the case.

   (b) Implement the example, but using a *Gaussian Basis Function* model instead.

      (i) First, try a single Gaussian basis function, set $\mu = \bar{(}x)$ (so the function is centered at the center of the data), and pick the value of s using cross-validation. Is this approach *globally* or *locally* sensitive to changes in x?

      (ii) Now, increase the number of basis functions you use to some fixed $M > 1$. Can you find sensible values of the μ_j? Remember the feature space is one-dimensional, so the j index refers only to the number of basis functions you choose to use; they are all functions of the same single feature x. Pick some μ_j yourself, then re-run cross-validation to pick s. Compare your answer to the previous question (when you used one basis function), and explain any differences.

      (iii) Can you simultaneously choose M, μ_j, and s, all using cross validation? Give it a try.

      (iv) *Optional*: Do you think you could do better if the training set was larger? If you're curious, try generating a few thousand points from the sinusoidal curve example from problem set 1, and re-running your procedure (this is less work than it sounds). Is CV more stable when there is more data?