Quiz 6

1 Using Lagrange multipliers for MAP estimation for a Categorical distribution

The Categorical distribution is defined as

$$\text{Cat}(y|\vec{\theta}) = \prod_{c=1}^{C} \theta_c^{1(y=c)},$$

for $y \in \{1, 2, \ldots, C\}$ where C is the number of labels and $C > 1$. We have observed N observations of Y, with N_k observation of each label. Suppose we decide to place a (unusual) zero-mean, identity-covariance Gaussian prior on $\vec{\theta}$, $p(\vec{\theta}) \propto \exp(-\theta^\top \theta)$.

You would like to find the MAP of $\vec{\theta}$.

1. Form the Lagrangian expression $\mathcal{L}(\vec{\theta}, \lambda)$.
2. Using your answer in Part 1., calculate the partial derivative with respect to each θ_k and λ.
3. Briefly describe how to use your answer to Part 2 to find the MAP for each θ_k. (You do not need to find an explicit solution.)

2 Optimizing Elastic Net using a LASSO solver

Define

$$J_1(w) = |y - Xw|^2 + \lambda_2 |w|^2 + \lambda_1 |w|_1$$

and

$$J_2(w) = |\tilde{y} - \tilde{X}w|^2 + c\lambda_1 |w|_1$$

where $c = (1 + \lambda_2)^{-\frac{1}{2}}$ and

$$\tilde{X} = c \begin{pmatrix} X \\ \sqrt{\lambda_2} I_d \end{pmatrix}, \; \tilde{y} = \begin{pmatrix} y \\ 0_{d\times 1} \end{pmatrix}$$

Show

$$J_1(w) = c(J_2(w))$$

i.e. $J_1(cw) = J_2(w)$ and hence that one can solve an elastic net problem using a lasso solver on modified data.