Assignment 6

Problem 1

The Categorical distribution is defined as

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

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.)

Problem 2

Consider a quadratic objective function on $\mathbf{R}^2$

$$f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2),$$

where $\gamma > 0$. We would like to apply the gradient descent method with exact line search, starting at the point $\vec{x}^{(0)} = (\gamma, 1)$.

1. Derive the exact line search update.
2. Suppose $\gamma = 10$, what is the value of $\vec{x}^{(3)}$?

Problem 3

In many classification problems one has the option either of assigning x to class j or, if you are too uncertain, of choosing a “reject” option. If the cost for rejects is less than the cost of falsely classifying the object, it may be the optimal action. Let α_i mean you choose action i, for $i = 1 : C + 1$, where C is the number of classes and $C + 1$ is the reject action. Let $Y = j$ be the true (but unknown) state of nature. Define the loss function as follows

$$\lambda(\alpha_i|Y = j) = \begin{cases} 0 & \text{if } i = j \text{ and } i, j \in \{1, \ldots, C\} \\ \lambda_r & \text{if } i = C + 1 \\ \lambda_s, & \text{otherwise} \end{cases}$$

In other words, you incur 0 loss if you correctly classify, you incur λ_r loss (cost) if you choose the reject option, and you incur λ_s loss (cost) if you make a substitution error (misclassification). Both λ_r and λ_s are positive numbers.

1. Show that when we decide to choose a class (and not reject), we always pick the most probable one.
2. Show that the minimum risk is obtained if we decide to pick the most probable class $j_{max} = \arg\max_j p(Y = j|x)$ and if $p(Y = j_{max}|x) \geq 1 - \frac{\lambda_r}{\lambda_s}$; otherwise we decide to reject.
3. Describe qualitatively what happens as λ_r/λ_s is increased from 0 to 1 (i.e., the relative cost of rejection increases).

Problem 4

Please write one thing from this course you found confusing, a topic you would like to hear more about, or something you found particularly interesting.