Multinomials and Multinomial Logistic Regression

Submit a PDF of your answers to Canvas

1. Consider a categorical random variable that takes one of K values according to a distribution $\boldsymbol{p} = [p_1 \;\; \cdots \;\; p_K]^T$. The multinomial distribution represents the number of outcomes of each category in n trials, extending the binomial distribution to $K > 2$. The pmf of the multinomial distribution with multinomial parameter $\boldsymbol{p}$ is given as

$$\frac{n!}{k_1!k_2!\cdots k_K!} p_1^{k_1} \cdots p_K^{k_K}$$

where k_i represents the number of outcomes of the ith category. Find and simplify an expression for Kullback–Leibler divergence between two multinomial distributions with multinomial parameters $\boldsymbol{p}$ and $\boldsymbol{q}$ with n trials. Relate your answer to the KL divergence between the underlying categorical random variables: $D(\boldsymbol{p}||\boldsymbol{q}) = \sum_i p_i \log \frac{p_i}{q_i}$.

2. In this problem we will implement two classifiers on the MNIST dataset.

a) Implement multinomial logistic regression for classification of MNIST images with the help of existing Python libraries (such as `sklearn.linear_model.LogisticRegression`). You may use the associated starter notebook if you choose. What is training error rate? What is test error rate?

b) Implement a single layer neural network using Python and keras (or another library, if you choose). Use the cross entropy loss function, and the softmax activation. You may use the starter notebook if you choose. What is training error rate on MNIST? What is test error rate on MNIST?

c) Which classifier performed better and why do you suspect that was the case? Which was faster to train?

3. (Optional) In order to show Shannon's Source Coding Theorem, we argued that

$$\log \binom{n}{\theta n} \approx nH_\theta$$

for large n and $\theta \in (0,1)$ and where $H_\theta = \theta \log_2 \frac{1}{\theta} + (1-\theta)\log_2 \frac{1}{1-\theta}$ is the binary entropy function. The argument required using Sterling's approximation: $n! \approx n^n$ for large n.

a) Find an actual upper and lower bound for the factorial function. *Hint:* It may be helpful to start by showing the following is true (with the help of a picture):

$$\int_1^n \log(x)\,dx \le \log(n!) \le \int_1^{n+1} \log(x)\,dx.$$

b) Use your bounds to show that

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \binom{n}{\theta n} = H_\theta.$$

It may be helpful to use the fact that $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.