

## 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  $\mathbf{p} = [p_1 \dots p_K]^T$ . 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  $\mathbf{p}$  is given as

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

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

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?