STAT3007/7007 Deep Learning, Tutorial 11
2022 Semester 2

1. (Autoencoders)

    (a) PCA can be viewed as a special kind of autoencoder. True or false? Justify your answer.

    (b) Sparse autoencoders aims to learn codes that better capture regularity in the data by encouraging sparsity in the code Describe how this is done using regularization, and give 3 example regularizers.

2. (KL-divergence) Recall that for two discrete distributions p and q defined on the same probability space, $KL(p||q)$ is defined as $\sum_x p(x) \ln \frac{p(x)}{q(x)}$.

    (a) Show that the KL-divergence is non-negative, that is, the KL-divergence has a lower bound of 0.

    (b) Does the KL-divergence have an upper bound? In particular, given any p defined on a probability space with at least 2 elements, is there an upper bound for $KL(p||q)$? Justify your answer.

    (c) The KL-divergence is often considered as a distance measure between two distributions. Show that the KL-divergence is not symmetric and does not satisfy the triangle inequality.

    (d) For two Bernoulli distributions $B(p)$ and $B(q)$, denote their KL-divergence by $KL(p,q) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q}$. Plot $KL(0.25, q)$ as a function of q, and also plot $KL(p, 0.25)$ as a function of p. Are the two functions convex? Can you generalize your observation and prove it?