Overfitting and Generalizability

1. You have been monitoring the accuracy of the weather predictions from Environment Canada for a few months, and calculate that the cross entropy of their predications for **Probability Of Precipitation** (**POP**) is 0.25. Hoping you can do better, you create a large neural network to predict the next day's POP based on a corpus of meteorological measurements. As a training set, you have 25 samples, (a_p, b_p), $p = 1, \ldots, 25$, where a_p is a vector of meteorological measurements, and b_p is the POP for the following day. You used those labelled samples to train your neural network using backpropagation. As a result of that training, your average cross entropy (CE) went from 1.45 (before training) to 0.00049 (after training).

   You are then given a test set of 25 more labelled samples to see how good your network performs. The average CE on these new samples is 0.58. You are given 25 more samples, and the average CE is 0.53.

   (a) Would you trust this model to predict the probability of precipitation tomorrow? Why or why not?

   (b) Briefly outline two methods that will probably improve your network's predictive abilities.

2. Download the jupyter notebook `ex05_torch.ipynb` and run it. It has a few suggestions for exercises (in the blue boxes), but I'm sure you can think of other ways to tinker with the code. Have fun!

3. If you have a neural network model, $y = f(x;\theta)$, where θ represents the weights, training the network is done by solving

$$\min_{\theta} E\big(f(X;\theta), T\big) \tag{1}$$

   where X and T are the inputs and targets (respectively) for the dataset. Using gradient descent, the update formula for the weights is

$$\theta \leftarrow \theta - \kappa \nabla_\theta E\big(f(X;\theta), T\big)\,. \tag{2}$$

   (a) Suppose you want to regularize the optimization process by seeking a set of weights for (1) that also minimizes the L_1 norm of the weights. In this context, the L_1 norm is simply the sum of the absolute values. Write down the cost function for this new, regularized optimization problem.

   (b) Write down the corresponding update formula for θ.