Homework 6
(due 02:00, Tuesday, November 20, 2018)

1. Write down the Lagrangian $L(w, b, \lambda)$ corresponding to the following optimization problem. Minimize

$$E(w) = \frac{1}{2} w^T w$$

subject to the constraints

$$y_i(w^T x_i + b) \geq 1, i \in \{1, \ldots, N\}.$$

Here, $x_i \in \mathbb{R}^D$ and $y_i \in \{-1, +1\}$. Then, eliminate (w, b) from the Lagrangian to obtain the "dual" objective function $W(\lambda)$. To achieve this, solve for (w, b) in terms of λ and plug the solution back into the objective function to obtain $W(\lambda)$. (Note that λ can be a set.) Are there any new constraints that emerge from performing this procedure? Give a qualitative explanation of what the objective function is trying to achieve.

2. Repeat the MNIST experiments conducted at `http://adventuresinmachinelearning.com/keras-tutorial-cnn-11-lines/`. Focus on the following. (i) Vary the convolution window size from 3×3 to 9×9, (ii) use two different nonlinearities (including RELU), (iii) try two different optimizers (SGD and Adam), (iv) pare the network down to just one convolution layer and one fully connected layer (no max pooling) and compare with the author's original network in terms of performance (training set error, test set error). For the author's network, show the images at the output of the CNN and max pooling layers and comment. Full points will be given only to training with a comprehensive cross product of parameter settings and comparisons. You must show the training curves w.r.t. epochs. Use a 50% split between training and testing taking care to see that all classes are roughly equally represented in training.