COMP527
Data Mining and Visualisation
Problem Set 2

Danushka Bollegala

Question 1 Let us consider the hinge loss $h(y) = \max(0, y)$. Given a train dataset $\mathcal{D} = \{(\boldsymbol{x}_i, t_i)\}_{i=1}^{N}$, we define the loss of classifying an instance (x_n, t_n) by $h(-t_n\boldsymbol{w}^\top\boldsymbol{x}_n)$. Here, $t_n \in \{1, -1\}$ is the target label of the instance $\boldsymbol{x}_n$. Answer the following questions about the derivation of the perceptron update rule.

A. Plot the hinge loss as a function of y.

B. Compute the differential $h'(y) = \frac{dh(y)}{dy}$.

C. Let us define the loss associated with a single instance to be $L(\boldsymbol{x}_n, t_n) = h(-t_n\boldsymbol{w}^\top\boldsymbol{x}_n)$. Show that this loss function reflects the *error-driven learning* approach on which perceptron is based.

D. Write the stochastic gradient descent rule for obtaining a new vector $\boldsymbol{w}^{(t+1)}$ from the current weight vector $\boldsymbol{w}^{(t)}$ after observing a train instance $(\boldsymbol{x}_n, t_n)$. Assume learning rate to be η.

E. Show that when $\eta = 1$ the update rule you derived in part D becomes the perceptron update rule.

F. How does regularization prevent overfitting?

G. Let us now add an ℓ_2 regularizer $||\boldsymbol{w}||^2 = \boldsymbol{w}^\top\boldsymbol{w}$ to our loss function to design the following objective function.

$$L(\boldsymbol{x}_n, t_n) = h(-t_n\boldsymbol{w}^\top\boldsymbol{x}_n) + \lambda||\boldsymbol{w}||2$$

Derive the perceptron update rule for this case.

H. Write the update rule for the logistic regression classifier.

I. Comparing part E and G, write the update rule for the regularised logistic regression.