Introduction to Machine Learning
Problem Set: Regularization, Logistic Regression

Summer 2021

1. (From Ethem Alpaydin's Introduction to Machine Learning) Consider the multi-class logistic regression with two classes using the softmax function.

   (a) Express the softmax outputs $g_0(z)$ and $g_1(z)$ in terms of $z_0 = w_0^T x$, and $z_1 = w_1^T x$.

   (b) Show that using two softmax outputs is equivalent to using one sigmoid output.
   Hint: if you write out $P(y = 0|x)$ and $P(y = 1|x)$ for the softmax function in terms of z_0 and z_1, and also write the sigmoid function output in terms of z, you can show that the two expressions are equivalent, for a particular relationship between z and z_0, z_1.

2. (By Prof. Sundeep Rangan) *Selecting a regularizer.* Suppose we fit a regularized least squares objective,

$$J(w) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \alpha\phi(w),$$

   where $\hat{y}_i$ is some prediction of y_i given the model parameters w. For each case below, suggest a possible regularization function $\phi(w)$, and briefly explain.

   There is no single correct answer. The answer may not necessarily be a "popular" regularization penalty you have already seen.

   (a) w should be sparse (i.e., only a few coefficients of w are nonzero).

   (b) the entries of w should be small on average.

   (c) negative coefficients are unlikely (but still possible), and very large negative coefficients are especially unlikely, but positive coefficients are not penalized.

   (d) each w_j (except for the first one) should be similar to the previous coefficient w_{j-1}.
   (Note: we are looking for a solution that achieves this very specifically, not just a solution that makes *all* the coefficients similar.)

3. *Handwritten digit classification.*

   Please refer to the homework notebook posted on the class site.