Exercises
Introduction to Machine Learning
FS 2018

Series 2, Mar 5th, 2018
(Model selection and Classification)

We will publish sample solutions on Friday, Mar 16th.

Problem 1 (Model selection and cross-validation):
Suppose we are given a noise-free set of points \(X = \{x_i\}_{i=1}^n \subset (-1,1), Y = \{\sin(x_i)\}\), which we want to fit with a polynomial, but we do not know which degree to choose. Suppose our candidate polynomial families are \(P_k = \{P_{2i+1}\}_{i=0}^k\), where \(P_{2i+1}\) denotes the family of polynomials with real-valued coefficients of maximum degree \(2i + 1\). We want to find the optimal hyperparameter value \(\hat{k} \in \{1, \ldots, k\}\).

Given a family of polynomials \(P_{2\ell+1}\) and a training set, suppose we have an oracle (i.e. an exact algorithm) that is able to find the polynomial \(\hat{p} \in P_{2\ell+1}\) with optimal coefficients with respect to the square loss objective
\[
\mathcal{L}(X, Y, p) = \sum_{i=1}^{n} (y_i - p(x_i))^2, \quad p \in P_{2\ell+1}.
\]

1. Show that when the optimization is performed on each family in \(P_k\), the lowest score is achieved when \(\hat{p} \in P_{2k+1} \setminus P_{2k-1}\) (i.e., \(\hat{p}\) will be of degree \(2k + 1\)).

2. What potential issue with using cross-validation does this demonstrate?

3. Suppose we add noise to the data, \(\tilde{Y} = \{y_i + \varepsilon_i\}_{i=1}^n\), where \(\varepsilon_i \sim \mathcal{N}(0, \sigma^2)\). For which values of \(\sigma^2\) will the result from part 1 hold with > 95% probability?

4. Suppose we widen the boundaries of \(X\) to \((-2\pi, 2\pi)\). Write a short script to simulate samples \((X_i, \tilde{Y}_i)\) with different values of \(\sigma_i^2\) and use 10-fold cross-validation to find corresponding optimal values \(\hat{k}_i\). How do \(\mathcal{L}(X_i, \tilde{Y}_i, \hat{p})\) and \(\hat{k}_i\) behave as \(k\) and \(\sigma^2\) increase?
Problem 2 (Classification):
Consider the data set plotted below,

Show that $\alpha = \frac{1}{||w||}$. How would $L_2$ regularization on $w$ affect the margin around $w^T x = 0$?