# Homework 3, ECE 6254, Spring 2014

Due: Wednesday Feb 26, at the beginning of class

**Problems:**

1. In this problem we will consider a simple learning scenario where $x \in \mathbb{R}$ and $y \in \mathbb{R}$ is given by $y = x^2$. Assume that the input variable $x$ is drawn uniformly on the interval $[-1, 1]$. Now suppose that we are given two independent observations of input-output pairs, i.e., our data set is given by $\mathcal{D} = \{(x_1, x_1^2), (x_2, x_2^2)\}$. In this problem we will consider two different approaches to fitting a line to this data set, and we will be interested in calculating the expected risk (in terms of squared error), the bias, and the variance.

   (a) To begin, suppose that our line is of the form $h(x) = b$, i.e., it is a constant function. We will fit the line by setting $h_{\mathcal{D}}(x) = (y_1 + y_2)/2 = (x_1^2 + x_2^2)/2$. For this case, analytically derive the average hypothesis
   $$\bar{h}(x) = \mathbb{E}_{\mathcal{D}}\left[h_{\mathcal{D}}(x)\right].$$

   (b) Next, analytically compute the bias, i.e.,
   $$\mathbb{E}_X\left[(\bar{h}(X) - X^2)^2\right].$$

   (c) Now compute (analytically) the variance
   $$\mathbb{E}_X\left[\mathbb{E}_{\mathcal{D}}\left[(h_{\mathcal{D}}(X) - \bar{h}(X))^2\right]\right].$$

   (d) Describe (in words) a simulation you could run in MATLAB to numerically estimate $\bar{h}(x)$, the bias, the variance, and the risk $\mathbb{E}_{\mathcal{D}}[R(h_{\mathcal{D}})]$ for this problem.

   (e) Run your simulation and report the results. Compare $\mathbb{E}_{\mathcal{D}}[R(h_{\mathcal{D}})]$ with the sum of the bias and variance computed analytically above.

   (f) Now suppose that instead we consider a line of the form $h(x) = ax + b$ which we fit by selecting the line the passes through our two observations. Modify the code from part (e) to estimate the new $\bar{h}(x)$, bias, variance, and risk. Comment on how the results change and offer an explanation for why they change in the way that they do.

2. In this problem we consider the scenario described in class, where $x$ is drawn uniformly on $[-1, 1]$ and $y = \sin(\pi x)$ and we are again given $n = 2$ training samples. Here we will consider an alternative approach to fitting a line to the data based on Tikhonov regularization. Specifically, we let
   $$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \qquad \boldsymbol{\theta} = \begin{bmatrix} b \\ a \end{bmatrix}.$$

   We will then consider Tikhonov regularized least squares estimators of the form
   $$\widehat{\boldsymbol{\theta}} = (\mathbf{A}^T\mathbf{A} + \boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1}\mathbf{A}^T\mathbf{y}. \tag{1}$$

(a) How should we set $\mathbf{\Gamma}$ to reduce this estimator to fitting a constant function (i.e., finding an $h(x)$ of the form $h(x) = b$)? (Hint: For the purposes of this problem, it is sufficient to set $\mathbf{\Gamma}$ in a way that just makes $a \approx 0$. To make $a = 0$ exactly requires setting $\mathbf{\Gamma}$ in a way that makes the matrix $\mathbf{A}^T\mathbf{A} + \mathbf{\Gamma}^T\mathbf{\Gamma}$ singular — but note that this does not mean that the regularized least-squares optimization problem cannot be solved, you must just use a different formula than the one given in (1).)

(b) How should we set $\mathbf{\Gamma}$ to reduce this estimator to fitting a line of the form $h(x) = ax + b$ that passes through the observed data points $(x_1, y_1)$ and $(x_2, y_2)$?

(c) Use the same approach as in the previous problem to numerically estimate the bias and variance for (at least approximations of) both of these estimators, and confirm that your estimates correspond to the numbers I provided in class.

(d) Play around and see if you can find a matrix $\mathbf{\Gamma}$ that results in a smaller risk than either of the two approaches we discussed in class. Report the $\mathbf{\Gamma}$ that gives you the best results.

3. In class we stated the result that a symmetric function $k(x, x')$ is an inner product kernel if and only if it is a positive semi-definite kernel. Show the "only if" part of this equivalence, i.e., show that an inner product kernel must be positive semi-definite. (The other direction is a bit more challenging.)

4. Find a feature map $\Phi$ that corresponds to the kernel

$$k(u, v) = (\mathbf{u}^T\mathbf{v} + 1)^3,$$

where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$.

5. Consider the feature map $\Phi : \mathbb{R}^2 \to \mathbb{R}^6$ defined by

$$\Phi(\mathbf{x}) = \left[1, x(1), x(2), x^2(1), x^2(2), x(1)x(2)\right]^T.$$

Using $\Phi$, find a $\mathbf{w} \in \mathbb{R}^6$ that parameterizes a hyperplane that represents the following decision boundaries.

(a) The parabola $(x(1) - 3)^2 + x(2) = 1$.

(b) The circle $(x(1) - 3)^2 + (x(2) - 2)^2 = 1$.

(c) The ellipse $2(x(1) - 3)^2 + (x(2) - 2)^2 = 1$.

(d) The hyperbola $(x(1) - 3)^2 - (x(2) - 4)^2 = 1$.

(e) The line $2x(1) + x(2) = 1$.