# Introduction to machine learning 89-511, Fall 2022 Home Assignment 1

Gal Chechik

Submission date: 2022-11-08

## Submit Instructions

Submit to the 'submit' system two files:
- 'knn.ipynb' with your solution to question no. 1
- 'solutions.pdf' with your solutions to questions no. 2-6

## 1. (30 pts) $k$ Nearest neighbor classification

See `colab.research.google.com/drive/1O4nq6bDOCGdfUcty9ohL7sIsk4Eb1Slf`

## 2. (15 pts) Polynomial regression

We showed that for a zero-order polynomial (namely, a constant $h_{\mathbf{w}}(x) = w_0$), the value that minimizes the mean squared error $Err(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(h_{\mathbf{w}}(x_i) - y_i)^2$ is the empirical *mean* of samples: $h_{\mathbf{w}}(x) = \frac{1}{n}\sum_{i=1}^{n} y_i$. Prove that for the case of zero-order polynomial with an absolute-value error

$$Err(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} |h_{\mathbf{w}}(x_i) - y_i| \quad ,$$

the optimal solution is the *median* of samples.

## 3. (10 pts) Computational complexity of $k$-NN

You are given a dataset of n labeled samples, where each input sample is a vector in a $d$-dimensional Euclidean space $x_1, \ldots, x_n \in R^d$. You wish to apply a $k$-NN algorithm using the Euclidean distance as the distance measure.

(a) What is the runtime complexity and memory complexity,in terms of $d$ and $n$, for training the classifier?

(b) What is the runtime complexity and the memory complexity, in terms of $d$ and $n$, for inferring the label of a new sample $x$?

## 4. (15 pts) Regularized polynomial regression

We derived in class the solution for a zero-degree polynomial regression. Consider the problem of regularized polynomial regression.

$$Err(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(h_w(x_i) - y_i)^2 + \lambda||\mathbf{w}||^2 \quad .$$

(a) Derive the solution for a polynomial of degree 0: $h_w(x) = w_0$ . Analyze the solution in the limit of $\lambda \to \infty$ and $\lambda \to 0$.

(b) Derive the solution for a polynomial of degree 1, $h_{\mathbf{w}}(x) = w_0 + w_1 x$, by computing the derivatives w.r.t. $w_0$ and $w_1$ and writing a system of two linear equations in $w_0$ and $w_1$. No need to solve the system. Analyze the solution in the limit of $\lambda \to \infty$ and $\lambda \to 0$.

## 5. (10 pts) PAC learning: Sample-Complexity Monotonicity

Let $\mathcal{H}$ be a hypothesis class for a binary classification task. Suppose that $\mathcal{H}$ is PAC learnable and its sample complexity is given by $N(\epsilon, \delta)$. Show that $N$ is monotonically non-increasing in each of its parameters. That is, show that given $\delta \in (0,1)$, and given $0 < \epsilon_1 \leq \epsilon_2 < 1$, we have that $N(\epsilon_1, \delta) \geq N(\epsilon_2, \delta)$. Similarly, show that given $\epsilon \in (0,1)$, and given $0 < \delta_1 \leq \delta_2 < 1$, we have that $N(\epsilon, \delta_1) \geq N(\epsilon, \delta_2)$.

## 6. (20 pts) PAC learnability of L2-balls around the origin

Given a real number $r > 0$, define the hypothesis $h_r : \mathbb{R}^d \to \{0, 1\}$ by:

$$h_r = \begin{cases} 1 & ||x||_2 < r \\ 0 & otherwise \end{cases} \tag{1}$$

Consider the hypothesis class $\mathcal{H} = \{h_r | r > 0\}$. Prove directly (without just using the fundamental theorem of PAC learning) that it is PAC learnable in the realizable case. Assume for simplicity that the marginal distribution of $X$ is continuous. How does the sample complexity depend on the dimension $d$ ? Explain.