Theoretical homework 1

Solutions may be made with latex or written & scanned. Solutions should be short and clea

September 23, 2017

1. The more parameters are employed by machine learning algorithm, the more it has a tendency to overfit. Indeed, overfitting means "flexibility"of the model towards each observation, that in turn means high "degree of freedom"(large number of parameters). Consider classification results of two methods: Linear classifier and K-Nearest Neighbour. In D-dimensional space linear classifiers have about D weight parameters, while kNN has a single one – the number of nearest neighbours. It is clear that despite having only one parameter, the decision boundary of kNN is more complex and flexible, as opposite to linear classifier solution. But that contradicts the valid argument about flexibility and the number of parameters! Why is this happening with kNN? Justify your answer.

2. Suppose $x \in \mathbb{R}^D$ is a feature vector. Consider transformation $f = \Sigma^{-1/2}(x - \mu)$, where $\mu = \mathbb{E}x$, $\Sigma = cov[x, x]$, $\Sigma^{1/2}$ satisfies the property that $\Sigma^{1/2}\left(\Sigma^{1/2}\right)^T = \Sigma$ and $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$. Prove that this transformation will give new feature vector f with:

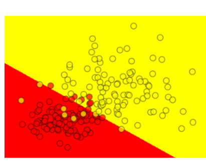
   (a) $\mathbb{E}f = \mathbf{0}$ (all zeroes vector)

   (b) $cov[f, f] = I$ (identity matrix)

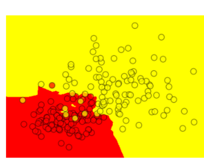
3. Suppose we need to estimate the center of distribution of real numbers $z_1, z_2, ...z_D$ in terms of their robustness to outliers (rare values which have are very distant from other values and have very low probability). Robust measure is a measure that does not change much when we add distant outliers to our training sample.

   (a) Is arithmetic average be robust to outliers? Why?

   (b) Is median robust to outliers? Why?



Decision boundary of the linear classifier



Decision boundary of kNN

4. Consider real numbers $z_1, z_2, ...z_N$. Find such constant approximation μ of these numbers, so that

   (a) $\sum_{n=1}^{N} (z_n - \mu)^2$ is minimized.

   (b) $\sum_{n=1}^{N} |z_n - \mu|$ is minimized.

5. Let $X \in \mathbb{R}^{NxD}$ be design matrix (objects x features). Consider series of optimization tasks:

   (a) a_1 is selected to maximize $\|Xa_1\|^2$ subject to $\langle a_1, a_1 \rangle = 1$

   (b) a_2 is selected to maximize $\|Xa_2\|^2$ subject to $\langle a_2, a_2 \rangle = 1$, $\langle a_2, a_1 \rangle = 0$

   (c) a_3 is selected to maximize $\|Xa_3\|^2$ subject to $\langle a_3, a_3 \rangle = 1$, $\langle a_3, a_1 \rangle = \langle a_3, a_2 \rangle = 0$
   etc.

   Prove that $a_1, a_2, a_3, ...$ are first, second, third,... largest eigenvectors (ordered by decreasing eigenvalue) of matrix $X^T X$.

   Hints:

   - recall that $\|x\| = \sqrt{\langle x, x \rangle}$, $\langle a, b \rangle = a^T b$, $\frac{\partial [a^T x]}{\partial x} = a$, $\frac{\partial [x^T x]}{\partial x} = 2x$, $\frac{\partial [x^T A x]}{\partial x} = 2Ax$ for any symmetric matrix A.
   - use method of Lagrange multipliers. To find dual coefficients, try multiplying by a_1^T, a_2^T, ...