EECE 5644 Introduction to Machine Learning and Pattern Recognition — June 18, 2023
Northeastern University, Summer I 2023 — Handout #2
Due: June 23

Problem Set 3

Problem 3.1 (Nearest neighbor predictors and Voronoi sets)

Nearest neighbor, k-nearest neighbor, and tree predictors are piecewise constant functions. This means that we can partition $\mathbf{R}^d$ into N regions, $\mathcal{R}_1, \ldots, \mathcal{R}_N$, and we have $g(x) = \hat{y}_k$ for all $x \in \mathcal{R}_k$. The regions don't overlap, except for their boundaries, and every point in $\mathbf{R}^d$ is in one of the regions. (We are not concerned with how these predictors work when a point is right on the boundary between two regions; in any case, whichever value the predictor takes on these boundaries makes no difference at all in practice.) In this problem we explore this idea.

(a) The set of points closer to one given point than another. Suppose u and v are given vectors in $\mathbf{R}^d$, with $u \neq v$. We define the set of all points in $\mathbf{R}^d$ that are closer to $u \in \mathbf{R}^d$ than $v \in \mathbf{R}^d$ as $\mathcal{S}(u, v) = \left\{x \in \mathbf{R}^d \mid \|x-u\|_2 \leq \|x-v\|_2\right\}$. Show that $\mathcal{S}(u, v)$ is a halfspace, which has the form $\mathcal{S}(u, v) = \left\{x \in \mathbf{R}^d \mid a^T x \leq b\right\}$. (You should say explicitly what the vector a is, and what the scalar b is.) The boundary of $\mathcal{S}(u, v)$ is a hyperplane, i.e., the set of points that satisfy $a^T x = b$. Sketch this for the specific example with $u = (1,1)$ and $v = (2,4)$. Show the points u and v, shade the set $\mathcal{S}(u, v)$, and indicate its boundary, which is a line (since a hyperplane in $\mathbf{R}^2$ is a line).

(b) Voronoi sets. Suppose $u^1, \ldots, u^m$ are given points in $\mathbf{R}^d$. Define $\mathcal{V}_i$, the set of points in $\mathbf{R}^d$ closer to x^i than the others (i.e., u^i is the nearest neighbor of all points in $\mathcal{V}_i$), for $i = 1, \ldots, m$. We can express these sets as

$$\mathcal{V}_i = \bigcap_{j \neq i} \mathcal{S}\left(u^i, u^j\right),$$

i.e., $\mathcal{V}_i$ is the intersection of the $m-1$ halfspaces $\mathcal{S}\left(u^i, u^j\right)$ for $j \neq i$. An intersection of halfspaces is called a polyhedron. The specific polyhedra $\mathcal{V}_i$ are called Voronoi sets or Voronoi regions associated with the collection of points $u^1, \ldots, u^m$. (Polyhedra is the plural of polyhedron.) They form a partition of $\mathbf{R}^d$ into a set of polyhedra, called the Voronoi partition of $\mathbf{R}^d$. Sketch the Voronoi regions for the collection of points $(1,1), (2,4), (5,3)$.

(c) Nearest neighbor predictor. Let g be the 1-nearest neighbor predictor for the data set $x^1, \ldots, x^n, y^1, \ldots, y^n$. Explain why g has the form $g(x) = y^k$ for $x \in \mathcal{V}_k$, where $\mathcal{V}_k$ is the Voronoi region associated with x^k. In other words, g is piecewise constant, with the same value inside each of the Voronoi regions associated with the vectors $x^i, i = 1, \ldots, n$.

(d) k-nearest neighbor predictor. Let $g(x)$ be the k-nearest neighbor predictor for the data set $x^1, \ldots, x^n, y^1, \ldots, y^n$. Explain why g is piecewise constant on a regions that are polyhedra. You can do this for the case $k = 2$, to make things simpler.

Hint. The set of points x for which x^q and x^l are its two nearest neighbors has the form

$$\mathcal{R}_{ql} = \left\{x \in \mathbf{R}^d \mid \|x-x^q\|_2 \leq \left\|x-x^j\right\|_2, \left\|x-x^l\right\|_2 \leq \left\|x-x^j\right\|_2, j \neq q \text{ or } l\right\}.$$

This is an intersection of $2(n-1)$ halfspaces. Hint. Drawing these sets in in 2D is encouraged. You can use such drawings as part of your explanation.

Problem 3.2 (Multiclass exponential loss)

For a K-class classification problem, consider the coding $Y = (Y_1, \dots, Y_K)^T$ with

$$Y_k = \begin{cases} 1, & \text{if } G = \mathcal{G}_k \\ -\frac{1}{K-1}, & \text{otherwise} \end{cases}$$

Let $f = (f_1, \dots, f_K)^T$ with $\sum_{k=1}^{K} f_k = 0$, and define

$$L(Y, f) = \exp\left(-\frac{1}{K} Y^T f\right)$$

(a) Using Lagrange multipliers, derive the population minimizer f^* of $E(Y, f)$, subject to the zero-sum constraint, and relate these to the class probabilities.

(b) Show that a multiclass boosting using this loss function leads to a reweighting algorithm similar to Adaboost

Problem 3.3 (Back propagation for cross-entropy loss function)

Derive the forward and backward propagation equations for the cross-entropy loss function.