Announcements

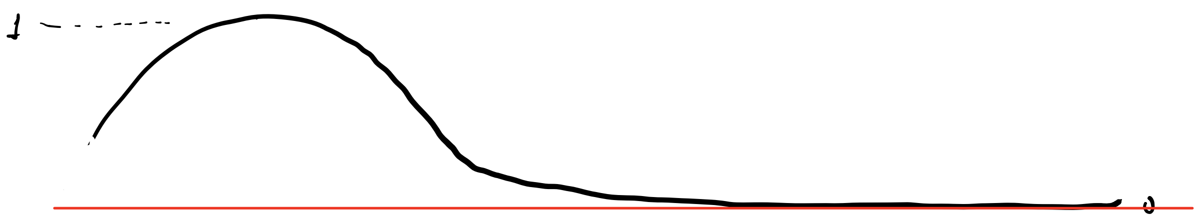
Quiz : Wednesday (linear algebra, optimization, probability)
HW2 : will be assigned this week

Outline

Recap Bayesian Classification
Theorem on optimal Bayes classifier

Inf (infimum) : greatest lower bound

Sup (supremum) : least upper bound



$\inf_{x} f(x) = 0$ $\qquad\qquad$ $\lim_{x \to \infty} f(x) = 0$

Guess the largest number $\qquad$ $X, Y \sim U[0,1]$ indep.
Observe X

Can't predict Y from X
Can predict $X > Y$

Bayes Classifiers

Given P_{XY} joint dist. of X, Y what is the best classifier possible?

$f : \mathbb{R}^d \to \{1, \dots, K\}$

Performance measure: probability of error (risk)

$$R(f) = P_{XY}\left(f(X) \neq Y\right)$$

$R^* =$ Bayes risk smallest risk of any classifier

$\pi_k = P_Y(Y=k)$ prior class probability

$g_k(x)$ pmf / pdf of $X \mid Y=k$

Define $\eta_k(x) \triangleq P_{Y \mid X=x}(Y=k \mid X=x)$

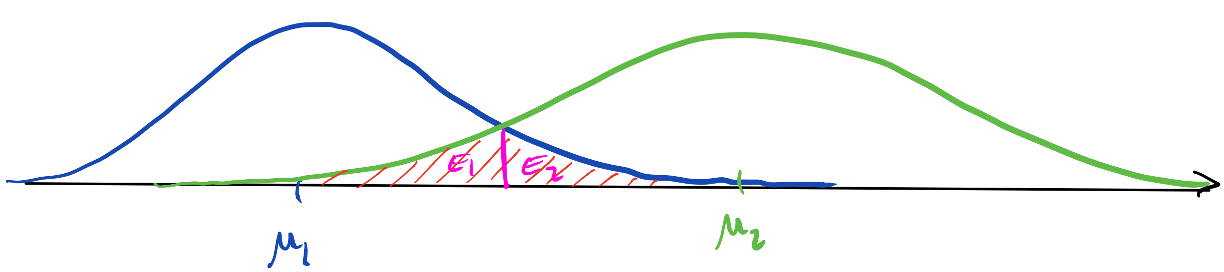
posterior class probability

$$\forall x : \sum_{k=1}^{K} \eta_k(x) = 1$$

Example: Suppose $X \mid Y$ is Gaussian. $Y \in \{-1,+1\}$

$$X \mid Y=-1 \sim N(\mu_-, \sigma_-^2)$$

$$X \mid Y=+1 \sim N(\mu_+, \sigma_+^2)$$



The error prob. is $R(f) = P_{XY}\left(f(X) \neq Y\right)$

Bayes decision function: $f^*(x) = \begin{cases} +1 & \text{if } \eta_1(x) > \frac{1}{2} \\ -1 & \text{otherwise} \end{cases}$

assuming classes are equally likely

$$\eta_1(x) = P(Y=1 \mid X=x) \quad \text{posterior +1 class probability}$$

$$= \frac{P(X=x \mid Y=1)\cdot P(Y=1)}{P(X=x)} \quad \text{(Bayes Rule)}$$

Theorem 1

$$f^*(x) = \arg\max_{k=1,\dots,K} \eta_k(x)$$

$$= \arg\max_{k=1,\dots,K} \pi_k \cdot g_k(x)$$

is a Bayes classifier.

Proof: Assume $X \mid Y=k$ is continuous. Let f be an arbitrary classifier

Decision Region: $\Gamma_k(f) = \{ x \mid f(x) = k \}$

$$1 - R(f) = P_{XY}(f(X) = Y) \quad \text{(prob. of correct classification)}$$

$$= \sum_{k=1}^{K} \pi_k \cdot \int_{\Gamma_k(f)} g_k(x)\, dx$$

$$\mathbb{1}_{\{x \in \Gamma_k(f)\}} = \begin{cases} 1 & \text{if } x \in \Gamma_k(f) \\ 0 & x \notin \Gamma_k(f) \end{cases}$$

$$= \sum_{k=1}^{K} \pi_k \cdot \int_{\mathbb{R}^d} g_k(x)\cdot \mathbb{1}_{\{x \in \Gamma_k(f)\}}\, dx$$

$$\int \sum^{K} \pi_k\, g_k(x) \cdot \mathbb{1} \qquad dx$$

$$= \int_{\mathbb{R}^d} \sum_{k=1} \pi_k \cdot g_k(x) \cdot \mathbb{1}_{\{x \in \Gamma_k(f)\}}$$

$\Gamma_1(f), \longrightarrow \Gamma_k(f)$ forms a partition of $\mathbb{R}^d$, i.e,
$\exists x$ $x \in \mathbb{R}^d$ belongs to only one $\Gamma_k(f)$.

to maximize $1 - R(f)$ choose

$$x \in \Gamma_k(f) \Leftrightarrow \pi_k \, g_k(x) \text{ is maximal.}$$

Bayes classifier $f^*(x) = \arg\max_k \pi_k \cdot g_k(x)$

Bayes Rule implies. $\eta_k(x) = \dfrac{\pi_k \cdot g_k(x)}{\sum_{\ell=1}^{K} \pi_\ell \cdot g_\ell(x)}$ $\longrightarrow$ independent of k. $\square$

$$\max_{\Omega_1, \Omega_2} \int f_1(x) \cdot \mathbb{1}_{\{x \in \Omega_1\}} dx + \int f_2(x) \cdot \mathbb{1}_{\{x \in \Omega_2\}} dx$$

$$\Rightarrow \Omega_1^* = \{ x \mid f_1(x) \geq f_2(x) \}$$

Nearest Neighbor Classification (NN)

Consider Binary classification problem

$$(x_1, y_1), \ldots, (x_n, y_n) \quad \text{where} \quad x_i \in \mathbb{R}^d, \; y_i \in \{-1, +1\}$$

x_i, y_i realizations of random pair X, Y

Given an unlabeled point x.

NN classifier : Assign x the same label as the closest training point x_i to x.

NN rule defines a partitioning of the feature space

Voronoi cells.