Machine Learning Problemset 2

Aimee Barciauskas

12 February 2016

7. Let the joint distribution of (X, Y) be such that X is uniform on the interval [0, 1], and for all $x \in [0,1], \eta(x) = x$. Determine the prior probabilities $\mathbb{P}\{Y = 0\}, \mathbb{P}\{Y = 1\}$ and the class-conditional densities $f(x|Y = 0)$ and $f(x|Y = 1)$. Calculate R^*, R_{1-NN}, R_{3-NN} (i.e., the Bayes risk and the asymptotic risk of the 1-, and 3-nearest neighbor rules).

R^*

$$R^* = \int_0^{\frac{1}{2}} \eta(x)dx + \int_{\frac{1}{2}}^1 (1 - \eta(x))dx$$

$$= \frac{x^2}{2}\Big|_0^{\frac{1}{2}} + [x - \frac{x^2}{2}]\Big|_{\frac{1}{2}}^1 = \frac{1}{4}$$

R_{1-NN}

From in-class calculation:

$$\begin{aligned} R_{1-NN} &= \eta(x)(1 - \eta(x)) + (1 - \eta(x)\eta(x) \\ &= 2\eta(x)(1 - \eta(x)) \\ &= \int [2\eta(x)(1 - \eta(x))]dx \end{aligned}$$

Substitute $x = \eta(x)$:

$$\begin{aligned} &= 2\int x(1 - x)dx \\ &= 2[\frac{1}{2}x^2 - \frac{1}{3}x^3]\Big|_0^1 \\ R_{1-NN} &= \frac{1}{3} \end{aligned}$$

R_{3-NN}

From in-class calculation we know:

$$R_{3-NN} = \mathbb{E}[\eta(x)(1 - \eta(x))] + 4\mathbb{E}[\eta(x)^2(1 - \eta(x))^2]$$

Following similar steps from R_{1-NN} we get:

$$\begin{aligned} &= \frac{1}{2}x^2 + \frac{3}{3}x^3 - \frac{8}{4}x^4 + \frac{4}{5}x^5\Big|_0^1 \\ R_{3-NN} &= \frac{3}{10} \end{aligned}$$

8. Let $X_1, ..., X_n$ be independent random variables taking values in $[0, 1]$. Denote $m = \mathbb{E}_{i=1} n X_i$. Prove that for any t:

$$\mathbb{P}\left\{ \sum_{i=1}^{n} X_i \geq t \right\} \leq \left(\frac{m}{t} \right)^t e^{t-m}$$

Hint: **Use Chernoff's bounding technique. Use the fact that by convexity of $e^{\lambda x}, e^{\lambda x} \leq xe^{\lambda} + (1 - x)$**

Start with the equality:

$$\mathbb{P}\{X \geq t\} = \mathbb{P}\{e^{\lambda X} \geq e^{\lambda t}\}$$

Using Chrenoff, we know this probability is less than or equal to:

$$\mathbb{P}\{e^{\lambda X} \geq e^{\lambda t}\} \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda t}}$$

We can set $\lambda = log(\frac{t}{m})$ because $t \geq m$ this will always be positive.

$$= \frac{\mathbb{E}[e^{\lambda X}]}{(\frac{t}{m})^t}$$

$$= (\frac{m}{t})^t \mathbb{E}[e^{\lambda X}]$$

Using the hint $e^{\lambda x}, e^{\lambda x} \leq xe^{\lambda} + (1 - x)$:

$$\leq (\frac{m}{t})^t \mathbb{E}[Xe^{\lambda} + 1 - X]$$

$$= (\frac{m}{t})^t [me^{\lambda} + 1 - m]$$

Substitute lambda:

$$= (\frac{m}{t})^t [t - m + 1]$$

The second term is always less than e^{t-m}:

$$\leq (\frac{m}{t})^t e^{t-m}$$

9. Let R_{k-NN} denote the asymptotic risk of the k-nearest neighbor classifier, where k is an odd positive integer. Use the expression of R_{k-NN} found in class to show that:

$$R_{k-NN} - R^* \leq \sup_{p \in [0, \frac{1}{2}]} (1 - 2p)\mathbb{P}\{Bin(k, p) > k/2\}$$

Proof:

$$R_{k-NN} - R^* = \mathbb{E}\Bigg[|2\eta(x)+1|\,\mathbb{P}\{Bin(k, min(\eta, 1-\eta) > \frac{k}{2}|X\} \Bigg]$$

$$\leq \mathbb{E}\Bigg[|2\eta(x)+1| \Bigg]\mathbb{P}\{Bin(k, min(\eta, 1-\eta) > \frac{k}{2}\}$$

$$= \mathbb{E}\Bigg[(1-2min(\eta, 1-\eta)) \Bigg]\mathbb{P}\{Bin(k, min(\eta, 1-\eta) > \frac{k}{2}\}$$

$$= 1 - 2\mathbb{E}[min(\eta, 1-\eta)]\mathbb{P}\{Bin(k, min(\eta, 1-\eta) > \frac{k}{2}\}$$

$$\leq \sup_{p\in[0,1/2]} (1-2p)\mathbb{P}\{Bin(k, min(\eta, 1-\eta) > \frac{k}{2}\}$$

Now we use Hoeffding's inequality to proof:

$$R_{k-NN} - R^* \leq \frac{1}{\sqrt{ke}}$$

For simplicity, we declare $B = Bin(k, p)$, and $p = min(\eta, 1-\eta)$ and we can reduce

$$\mathbb{P}\{Bin(k, min(\eta, 1-\eta) > \frac{k}{2}\} = \mathbb{P}\{B > \frac{k}{2}\}$$

$$\mathbb{P}\{B > \frac{k}{2}\} = \mathbb{P}\{B - kp > k(\frac{1}{2} - p)\}$$

Theorem 8.1, where S_n is a sum of random variables:

$$\mathbb{P}\{S_n - \mathbb{E}S_n \geq \epsilon\} \leq exp\{-2\epsilon^2 / \sum(b_i - a_i)^2\}$$

Since B is a sum of random variables and kp is it's expected value, the above expression can be expressed:

$$\mathbb{P}\{B - kp > k(\frac{1}{2} - p)\} \geq exp\{-2k(1-p) / \sum(b_i - a_i)^2\}$$

The sum in the exponent is always a sum of the difference between a binary variable, so it is a sum of k 1's.

$$\mathbb{P}\{B > \frac{k}{2}\} \geq exp\{-2k(1/2-p)^2\}$$

Replacing p with u for simplicity ($u = 1-2p$) the entire expression can be written as:

$$\sup_{p\in[0,\frac{1}{2}]} (1-2p)\mathbb{P}\{Bin(k,p) > k/2\} \leq \sup_{0\leq u\leq 1} ue^{-ku^2/2}$$

which is equivalent to

$$= \frac{1}{\sqrt{ke}}$$

10. (RADEMACHER AVERAGES) Let $\mathcal{A}$ be a bounded subset of R_n. Define the Rademacher Average:

$$R_n(A) = \mathbb{E} \sup_{a \in A} \frac{1}{n} \left| \sum_{i=1}^{n} \sigma_i a_i \right|$$

. where $\sigma_1, ..., \sigma_n$ are independent random variables with $\mathbb{P}\{\sigma_i = 1\} = \mathbb{P}\{\sigma_i = -1\} = \frac{1}{2}$. Prove the following "structural" results:

$$R_n(A \cup B) \leq R_n(A) + R_n(B)$$

Proof:

$$R_n(A \cup B) = \mathbb{E} \sup_{v \in A \cup B} \frac{1}{n} |\sigma_i v_i|$$

$$\leq \mathbb{E} \sup_{a \in A} \frac{1}{n} \left| \sum_{i=1}^{n} \sigma_i a_i \right| + \mathbb{E} \sup_{b \in B} \frac{1}{n} \left| \sum_{i=1}^{n} \sigma_i b_i \right|$$

$$R_n(c \bullet A) = |c| R_n A$$

Proof

$$R_n(c \bullet A) = \mathbb{E} \sup_{a \in c \bullet A} \frac{1}{n} \left| \sum_{i=1}^{n} \sigma_i c a_i \right|$$

$$= |a| \, \mathbb{E} \sup_{a \in \bullet A} \frac{1}{n} \left| \sum_{i=1}^{n} \sigma_i a_i \right|$$

$$R_n(A \oplus B) = R_n(A) + R_n(B)$$

Proof

$$R_n(A \oplus B) = \mathbb{E} \sup_{v \in A+B} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i v_i \right|$$

$$\mathbb{E} \sup_{a \in A, b \in B} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i (a_i + b_i) \right|$$

$$\leq \mathbb{E} \sup_{a \in A} \frac{1}{n} |\sigma_i a_i| + \mathbb{E} \sup_{b \in B} \frac{1}{n} |\sigma_i b_i|$$

$$= R_n(A) + R_n(B)$$

If $absconv(A) = \{\sum_{j=1}^{N} c_j a^j : N \in \mathbb{N}, \sum_{j=1}^{N} |c_j| \leq 1, a^j \in A\}$ **is the absolute convex hull of A, then:**

The absolute convex hull of A is the union of all sets:

$$c_1 A_1 + ... + c_N A_N = \{c_1 a_1 + ... + c_N a_N : a_1, ..., a_N \in A\}$$

then, the rademacher average of a given set:

$$R_n(c_1 A_1 + ... + c_N A_N) = \sum_{j=1}^{n} |c_j| \, R_n(A) \leq R_n(A)$$

So for all N choices of c_j, the Rademacher average of absolute convex hull of the set A is less than or equal to the Rademacher average of A, so the Rademacher average of A is equal to the Rademacher average of A.

11. A half plane is a set of the form $H_{a,b,c} = \{(x, y) \in \mathbb{R}^2 : ax + by \geq c\}$ for some real numbers a, b, c. Determine the n-th shatter coefficient of the classes:

$$\mathcal{A}_0 = \{H_{a,b,0} : a, b \in \mathbb{R}\} \text{ and } \mathcal{A}_0 = \{H_{a,b,c} : a, b, c \in \mathbb{R}\}$$

In the first case, $\mathcal{A}_0 = \{H_{a,b,0} : a, b \in \mathbb{R}\}$, a half plane must pass through the origin so it can only rotate around the origin. So n points can be subset into $n + 1$ different ways. So the n-th shatter coefficient is $n + 1$.

Corollary 13.1 defines the shatter coefficient for the class of all half-spaces. Adapting the equation for the class of half-spaces in $\mathbb{R}^2$ gives:

$$s(\mathcal{A}, n) = 2 \sum_{i=0}^{2} \binom{n-1}{i}$$

Expanding the sum we find the n-th shatter coefficient for the class of half-spaces in $\mathbb{R}^2$:

$$s(\mathcal{A}, n) = 2n + (n - 1)(n - 2)$$