Machine Learning Problem Sets

Aimee Barciauskas

29 March 2016

Set 1

Problem 1: Determine the Bayes Risk using class-conditional probabilities

$R^* = min(\eta(x), 1 - \eta(x))$

- Use: $\eta(x) = \frac{\mathbb{P}\{Y=1\}f_1(x)}{f(x)}$, where $f_1(x)$ is the conditional distribution of x given $Y = 1$
- Substitute prior probabilities, remove normalizing $f(x)$

Problem 2: Determine the Bayes Classifier using class-conditional probabilities

$$g^*(x) = \begin{cases} 1 & \text{if } q_1 f_1 > q_0 f_0 \\ 0 & \text{otherwise} \end{cases}$$

- Substituted f_1 and f_0 directly, reduce to find when the inequality holds

Problem 3 and 4: Determine a function which is the optimal classifier for a given loss function

- The function minimizes the expected loss: $\mathbb{E}[\ell(y, y')] = \int \ell(y, y')p(y|x)dy$
  - Take derivative wrt to function y', set equal to 0 and solve for y'
- Show that for any alternative function, the expected loss is greater than or equal to the loss of the function.

Problem 5: Show the probability the distance of the k-nearest neighbors is 0 goes to 1 as n goes to infinity.

- Show the probability that the distance is greater than epsilon goes to 0:
  1. Take the expected value of the probability wrt X
  2. Replace the probability expression with it's binomial expression: the probability that n samples are not in the ball is less than k
  3. This is less than or equal to the unconditional probability
  4. Solve the probability using the expression for the probability of a binomial:

  $\mathbb{P}\{Bin(n,q) < k\} = \sum_{i=1}^{k-1} nchooseiq^k(1-q)^{n-k}$

  5. This expression goes to zero with $n \to \infty$

Problem 6: Show $\mathbb{E}[R(g)] > \frac{1}{4}$ even when $R^* = 0$

- For 1-NN: Show the probability that nothing falls in the X_i's "bucket" is greater than $\frac{1}{4}$
  - e.g. $\mathbb{P}\{Bin(n, \frac{1}{m}\} \geq 2$

Set 2

Problem 7: Calculate R^*, R_{1-NN}, R_{3-NN}

- Expand R as a function of $\eta(x)$ and take expected value (e.g. integrate over possible values of $eta(x)$)

Problem 8: Show that the probability a sum of random indpendent variables taking values in $[0,1]$ is greater than t is bounded by a complicated function of it's expected value and t

1. Use Chernoff to rewrite in terms of expectations and a product over X_i's
2. Use convexity of $e^{\lambda x}$ to take x out of exponent
3. Bring through expected value of X_i
4. Use $1 + x \leq e^x$ to put expected value back in exponent, now can use sum of expected value in exponent (which will just be the expected value)
5. This will be some exponential function, which will be minimized when the exponent is minimized. So take derivate and set equal to lambda. Put lambda back in and reduce.

Problem 9: Show the deviation from R^* of $R(g)$

1. Restate R in terms of the expected risk, i.e.: probability of each type of error by the probability of each class $\eta(x)$, $1 - \eta(x)$
2. Is it possible to simplify to an R^* term? Now we have an expression for the deviance from R^*
3. Use Hoeffding to replace probabilty a random variable deviates from it's expected value by more than some expression, say t, is $\leq e^{-2nt^2}$

Problem 10: Prove "structural" results of some sets (e.g. Rademacher averages)

- Rewrite structural result in terms of full expression, note if it's an inquality and there is some supremum of infinum involved it could be trivial

Problem 11: Determine the n-th shatter coefficient

$s(\mathcal{A}, n) = max_{(z_1,...,z_n) \in \{\mathcal{R}^d\}} N_{\mathcal{A}}(z_1, ..., z_n)$ where N is the number of different sets which in union compose $\mathcal{A}$

The shatter coefficient is the maximal number of different subsets of n points that can be picked out by the class of sets $\mathcal{A}$