CS485/685
Lecture 21: March 20, 2012

Nearest neighbor analysis

[BDSS] Chapter 9

Nearest Neighbor Recap

Analysis Challenge

- How can we analyze nearest neighbor rules?
  - They don't have a well defined hypothesis space
    - # of hypotheses grows with the amount of data
    - Can't determine VC dimension
    - Non-parametric method
  - What is the learning bias?
    - No prior, no penalty term, no regularization…
    - Does this contradict the no-free lunch theorem?

Learning Bias

- Let η be the underlying stochastic labeling function:

$$\eta(x) = \Pr(Y = 1 | X = x)$$

- Nearest neighbor works well when labeling function η is ***c*-Lipschitz** (smooth)

$$\exists c > 0, \forall x, x' \; |\eta(x) - \eta(x')| \leq c||x - x'||_2$$

- When two inputs are near each other, their labeling distributions are similar

Generalization Bound

- **Lemma:** Let D be a distribution over $\Re \times \{0,1\}$ and assume η is c-Lipschitz. Let S be a sample of size N and h_S be its corresponding 1-NN rule. Then

$$E_S[L_D(h_S)] \leq 2L_D(h^*) + c\ E_{X,S}\left[\left|\left|X - x_{\pi_1(X)}\right|\right|_2\right]$$

where h^* is the hypothesis with minimum loss
$\pi_1(X)$ is the nearest neighbor of X in S

Bayes Optimal Classifier

- Suppose you know the underlying distribution D and the underlying labeling function η
- Then the hypothesis with minimum loss (a.k.a. Bayes Optimal Classifier) is

$$h^* = argmin_h\ L_D(h)$$

$$h^*(x) = \begin{cases} 1 & \eta(x) > 0.5 \\ 0 & \eta(x) \leq 0.5 \end{cases}$$

- The minimum loss is:

$$L_D(h^*) = E_X[\min\{\eta(X), 1 - \eta(X)\}]$$
$$\geq E_X\left[\eta(X)\big(1 - \eta(X)\big)\right]$$

Proof of Generalization Bound

- Consider the probability of sampling different labels for x and x'.

$$\Pr_{Y\sim\eta(x),Y'\sim\eta(x')}[Y \neq Y']$$
$$= \eta(x')\big(1-\eta(x)\big) + \big(1-\eta(x')\big)\eta(x)$$
$$= \eta(x') - 2\eta(x)\eta(x') + \eta(x)$$
$$= 2\eta(x)\big(1-\eta(x)\big) + \eta(x') - \eta(x)$$
$$\leq 2\eta(x)\big(1-\eta(x)\big) + c\big||x-x'|\big|_2 \text{ (c-Lipschitz property)}$$

Proof of Generalization Bound

- Consider the expected error of the 1-NN rule

$$E_S[L_D(h_S)]$$
$$= E_{S,(X,Y)}[Y \neq Y_{\pi_1(X)}] \text{ (by definition)}$$
$$\leq E_{S,(X,Y)}\left[2\eta(X)\big(1-\eta(X)\big) + c\left||X - x_{\pi_1(X)}|\right|_2\right] \text{ (prev slide)}$$
$$= 2E_X\big[\eta(X)\big(1-\eta(X)\big)\big] + cE_{S,X}\left[\left||X - x_{\pi_1(X)}|\right|_2\right]$$
$$\leq 2L_D(h^*) + cE_{S,X}\left[\left||X - x_{\pi_1(X)}|\right|_2\right] \text{ (Bayes optimal rule)}$$

1-NN Generalization Bound

- In the limit
  i.e. as $N \to \infty$ then $\left|\left|X - x_{\pi_1(X)}\right|\right|_2 \to 0$
  the error of the 1-NN rule is at most twice the minimum error.

- When N is finite, how do we bound $\left|\left|X - x_{\pi_1(X)}\right|\right|_2$?

Nearest neighbor distance bound

- Lemma: For a sample S of N inputs in $[0,1]^d$, we have that

$$E_{X,S}\left[\left|\left|X - x_{\pi_1(X)}\right|\right|_2\right] \leq 4\sqrt{d}N^{-\frac{1}{d+1}}$$

Cover Set

- Consider a set of hypercubes of length ϵ that covers the input space $[0,1]^d$

Distance Bound

- Consider two cases:

1. There is a neighbour in **same** hypercube

$$\left|\left|X - x_{\pi_1(X)}\right|\right|_2 \leq \epsilon\sqrt{d}$$

2. All neighbours in **different** hypercubes

$$\left|\left|X - x_{\pi_1(X)}\right|\right|_2 \leq \sqrt{d}$$

Bound Probability of different hypercubes

- **Lemma:** Let $C_1, C_2, \ldots, C_r$ be r hypercubes and S be a sequence of N input points sampled i.i.d. from D_X. Then

$$E_S\left[\Pr\left[\bigcup_{i:C_i\cap S=\emptyset} C_i\right]\right] \le \frac{r}{Ne}$$

Proof

$$\begin{aligned}
&E_S\left[\Pr\left[\cup_{i:C_i\cap S=\emptyset} C_i\right]\right] \\
&\quad \le E_S\left[\textstyle\sum_{i:C_i\cap S=\emptyset} \Pr(C_i)\right] \\
&\quad = \textstyle\sum_{i=1}^{r} \Pr(C_i)\, E_S[\delta(C_i \cap S = \emptyset)] \\
&\quad = \textstyle\sum_{i=1}^{r} \Pr(C_i)\, \Pr[C_i \cap S = \emptyset] \\
&\quad = \textstyle\sum_{i=1}^{r} \Pr(C_i)\, (1 - \Pr(C_i))^N \\
&\quad \le \textstyle\sum_{i=1}^{r} \Pr(C_i)\, e^{-\Pr(C_i)N} \\
&\quad \le \frac{r}{Ne}
\end{aligned}$$

Expected Distance Bound

$E_{X,S}\left[\left\|X - x_{\pi_1(X)}\right\|_2\right]$

$\leq E_S\Big[\underbrace{\Pr\big[\cup_{i:C_i \cap S = \emptyset} C_i\big]}_{\leq \frac{r}{Ne}} \sqrt{d} + \underbrace{\Pr\big[\cup_{i:C_i \cap S \neq \emptyset} C_i\big]}_{\leq 1}\ \epsilon\sqrt{d}\Big]$

$\leq \sqrt{d}\left(\frac{r}{Ne} + \epsilon\right)$

…

$\leq 4\sqrt{d}N^{-\frac{1}{d+1}}$

1-NN Generalization Bound

Theorem: Let h_S denote the result of applying the 1-NN rule to a sample S. Then

$$E_S[L_D(h_S)] \leq 2L_D(h^*) + 4c\sqrt{d}N^{-\frac{1}{d+1}}$$

Proof: combine previous lemmas