STATS 305B: Methods for Applied Statistics I **Winter 2016/17**

Lecture 19: February 24

Lecturer: Jonathan Taylor *Scribes: Kenneth Tay*

19.1 Smoothing and Penalized Methods

In the logistic regression setting, we were interested in modeling $\pi(x) = P(Y = 1 \mid X = x)$. Logistic regression models $\pi(x)$ as

$$\text{logit}(\pi(x)) = x^T \beta.$$

There are other ways to estimate $\pi(x)$.

19.1.1 Smoothing Kernels

Let K be some function which we call the **kernel function**. (A typical choice for the kernel function is Gaussian.) Given some bandwidth parameter λ, we can derive an estimate of $\pi(x)$:

$$\hat{\pi}_\lambda(x) := \frac{\displaystyle\sum_{i=1}^n y_i K\left(\frac{x_i - x}{\lambda}\right)}{\displaystyle\sum_{i=1}^n K\left(\frac{x_i - x}{\lambda}\right)}.$$

We can think of the $K\left(\dfrac{x_i - x}{\lambda}\right)$ terms as weights on the y_i's. Intuitively, we put more weight on y_i if x_i is "closer" to x.

Pros:

- No enforced structure, hence allows for very flexible models.
- No optimization step required.

Cons:

- As $p = \dim X$ grows, this method suffers quickly from the curse of dimensionality.
- The bandwidth parameter λ needs to be chosen. (A typical way to do this is through cross validation.)

Note: In the binary context, $\pi(x) = P(Y = 1 \mid X = x) = \mathbb{E}[Y \mid X = x]$. This method can be extended easily to estimate $\mathbb{E}[Y \mid X = x]$ where Y is real-valued.

19.1.2 k-Nearest Neighbors

Here, our estimate of $\pi(x)$ is

$$\hat{\pi}_k(x) := \text{mean of } x\text{'s } k \text{ nearest neighbors.}$$

We can also estimate $y(x)$ as

$$\hat{y}_k(x) := \begin{cases} 1 & \text{if } \hat{\pi}_k(x) \geq 0.5, \\ 0 & \text{otherwise.} \end{cases}$$

k-Nearest Neighbors can be thought of as a type of kernel smoothing, where the kernel function is $K(x) = 1_{B(r)}(x)$, where $B(r)$ is a ball of radius r centered at x, with r adaptively chosen so that there are k neighbors in the ball.

k-Nearest Neighbors has the same pros and cons as smoothing kernels.

19.1.3 Generalized Additive Models

We can think of logistic regression as

$$\text{logit}(\pi(x)) = \sum_{j=1}^{p} x_j \beta_j = \sum_{j=1}^{p} f_j(x_j),$$

where $f_j(x_j) = x_j \beta_j$. In principle, we could have more complicated functions for f_j, e.g. $f_j = \sum_k a_{jk} h_{jk}$. These are called **Generalized Additive Models**.

One example of the h_{jk}'s would be spline functions on the support of X_j.