Fundamentals of Inference and Learning

Homework 2

2 Statistical Learning with Nearest - Neighbors

$\mathbb{P}(\underline{X}, y)$

- $\underline{X} \in \mathbb{R}^d$, $y \in \mathbb{R}$
- $Var[y] = \sigma^2$

1. $\mathbb{P}(\underline{X}, y)$ is known

$$\mathbb{E}_{\underline{X},y}[(f(\underline{X}) - \hat{y})^2] \text{ , with } f_{Bayes}(X) = \mathbb{E}[y|\underline{X}]$$

minimizing: $\mathbb{E}_{\underline{X},y}[(f(\underline{X}) - y)^2] = \mathbb{E}_X[\mathbb{E}_{y|X}[(y|X - \hat{y})^2]]$

we therefore need to minimize $\mathbb{E}_{y|X}[(y|X - \hat{y})^2]$

$$\frac{\partial}{\partial q} \mathbb{E}_{y|X}[(y|X - \hat{y})^2] = 2\mathbb{E}_{y|X}[y|X - \hat{y}] = 0$$

$$\Leftrightarrow \hat{y} = \mathbb{E}_{y|X}[y|X] = f_{Bayes}(X)$$

2. $f_{kNN}(\underline{X}) = \frac{1}{k} \sum_{i \in N_k(X)} y_i$

the excess risk $\Delta R = R_{kNN} - R_{Bayes} = \mathbb{E}_{X,y,y_i,x_i}[(y - f_{kNN}(X))^2 - (y - f_{Bayes}(X))^2]$

$$= \mathbb{E}_X \mathbb{E}_{y|X,x_i,y_i}[(y|X - \frac{1}{k}\sum_i y_i)^2 - \underbrace{(y|X - \mathbb{E}_{y|X}[y|X])^2}_{= Var(y|X)}]$$

$$= \mathbb{E}_X[\mathbb{E}_{y|X,x_i,y_i}[(y|X)^2 + \frac{1}{k^2}\sum_{i,j \in N_k(X)} y_i y_j - \frac{2y|X}{k}\sum_{i \in N_k(X)} y_i] - Var(y|X)]$$

$$= \mathbb{E}_X[\mathbb{E}_{y|X}[(y|X)^2] + \frac{1}{k^2}\sum_{i,j} \mathbb{E}_{y_i}[y_i]\mathbb{E}_{y_j}[y_j] - \frac{2}{k}\mathbb{E}_{y|X}[y|X]\sum_i \mathbb{E}_{y_i}[y_i] - Var(y|X)]$$

$$= \mathbb{E}_X[\mathbb{E}_{y|X}[(y|X)^2] + \frac{1}{k^2}\sum_{i \neq j} \mathbb{E}_{y_i}[y_i]\mathbb{E}_{y_j}[y_j] + \frac{1}{k^2}\sum_{i=j} \mathbb{E}_{y_i}[y_i^2] - \frac{2}{k}\mathbb{E}_{y|X}[y|X]\sum_i \mathbb{E}_{y_i}[y_i]$$

$$- Var(y|X)]$$

But : - $y_i = y|x_i \Rightarrow \mathbb{E}(y_i) = f_{Bayes}(x_i)$

- $\mathbb{E}[y_i^2] = \Delta(x_i) + f_{Bayes}(x_i)$

- $\mathbb{E}[y|X] = f_{Bayes}(X)$

- $\mathbb{E}[(y|X)^2] = \Delta(X) + f^2_{Bayes}(X)$

Then $\mathbb{E}_X[\mathbb{E}_{y|x}[(y|X)^2] + \frac{1}{k^2}\sum_{i\neq j}\mathbb{E}_{y_i}[y_i]\mathbb{E}_{y_i}[y_j] + \frac{1}{k^2}\sum_{i=j}\mathbb{E}_{y_i}[y_i^2] - \frac{2}{k}\mathbb{E}_{y|X}[y|X]\sum_i \mathbb{E}_{y_i}[y_i]$

$- var(y|X)]$

$= \mathbb{E}_X[\cancel{var(y|X)} + f^2_{Bayes}(X) + \frac{1}{k^2}\sum_{i\neq j} f_{Bayes}(X_i)f_{Bayes}(X_j) + \frac{1}{k^2}\sum_{i=j} var(x_i) + f_{Bayes}(X_i)$

$- \frac{2}{k} f_{Bayes}(X)\sum_{i\in N_k(X)} f_{Bayes}(X_i) \quad - \cancel{var(y|X)}$

$= \mathbb{E}_X[\frac{1}{k}var(y|X) + f^2_{Bayes}(X) + \frac{1}{k^2}\sum_{i\neq j} f_{Bayes}(X_i)f_{Bayes}(X_j) + \frac{1}{k^2}\sum_{i=j} f_{Bayes}(X_i)^2$

$- \frac{2}{k} f_{Bayes}(X)\sum_{i\in N_k(x)} f_{Bayes}(X_i)]$

$= \mathbb{E}_X[\frac{1}{k}var(y|X) + (f_{Bayes}(X) - \frac{1}{k}\sum_{i\in N_k(X)} f_{Bayes}(x_i))^2]$

$= \frac{\sigma^2}{k} + \mathbb{E}_X(\frac{1}{k}\sum_{i\in N_k(X)} f_{Bayes}(X_i) - f_{Bayes}(X))^2]$

$= b^2 + v$

3. $f_{Bayes}(X)$ is L-lipschitz : $|f(X) - f(y)| \leq L\|X - y\|_2$

$\Rightarrow |f_{Bayes}(X) - f_{Bayes}(y)| \leq L\|X - y\|_2$

$* \; (\frac{1}{k}\sum_{i\in N_k(X)} f_{Bayes}(x_i) - f_{Bayes}(X))^2 \leq \frac{1}{k^2} k \sum_{i\in N_k(X)} (f_{Bayes}(X_i) - f_{Bayes}(X))^2$

$\Delta R \leq \frac{\sigma^2}{k} + \frac{1}{k}\sum_{i\in N_k(X)} (f_{Bayes}(X_i) - f_{Bayes}(X))^2$

$\leq \frac{\sigma^2}{k} + \frac{L^2}{k}\sum_{i\in N_k(X)} \|x_i - X\|_2^2$

$$\|x_i - x\|_2^2 \approx \ell^2 = \left(\frac{k}{n}\right)^{2/d}$$

$$\Rightarrow \Delta R \leq \frac{\sigma^2}{k} + L^2 \left(\frac{k}{n}\right)^{2/d}$$

5) $$\frac{\partial \Delta R}{\partial k} = -\frac{\sigma^2}{k^2} + \frac{L^2}{n^{2/d}} \cdot \frac{2}{d} k^{\underbrace{\frac{2}{d}-1}_{=\frac{2-d}{d}}}$$

$$= -\frac{\sigma^2}{k^2} + \frac{2L^2}{d\, n^{2/d}} \cdot k^{\frac{2-d}{d}} = 0$$

$$\Leftrightarrow \sigma^2 = \frac{2L^2}{d\, n^{2/d}} \cdot k^{\frac{2-d}{d} + \frac{2d}{d}} = \frac{2L^2}{d n^{2/d}} \cdot k^{\frac{2+d}{d}}$$

$$\Leftrightarrow k_{opt} = \left(\frac{d\, n^{2/d} \sigma^2}{2L^2}\right)^{d/2+d}$$

6) $$\Delta R \leq \frac{\sigma^2}{k} + L^2\left(\frac{k}{n}\right)^{2/d} \Big|_{k_{opt}}$$

$$= \sigma^2 \cdot \left(\frac{2L^2}{d n^{2/d} \cdot \sigma^2}\right)^{d/2+d} + L^2 \frac{\left(\frac{d\, n^{2/d} \sigma^2}{2L^2}\right)^{d/2+d \,\cdot\, 2/d}}{n^{2/d}}$$

$$= \sigma^2 \left(\frac{2L^2}{d n^{2/d} \cdot \sigma^2}\right)^{d/2+d} + L^2 \frac{\left(\frac{d\, n^{2/d} \sigma^2}{2L^2}\right)^{2/2+d}}{n^{2/d}}$$

Setting $L = \sigma = 1$

$$\Delta R \leq \left(\frac{2}{d n^{2/d}}\right)^{d/2+d} + \frac{\left(\frac{d n^{2/d}}{2}\right)^{2/2+d}}{n^{2/d}}$$

$$= \left(\frac{d}{2}\right)^{-d/2+d} \cdot n^{-\frac{2}{\cancel{d}} \cdot \frac{\cancel{d}}{2+d}} + n^{-\frac{2}{d}} \cdot \left(\frac{d}{2}\right)^{\frac{2}{2+d}} \cdot n^{\frac{2}{d} \cdot \frac{2}{2+d}}$$

$$= n^{-\frac{2}{2+d}} \left(\frac{d}{2}\right)^{-\frac{d}{2+d}} + n^{\underbrace{-\frac{2}{d}\left(1 - \frac{2}{2+d}\right)}_{= -\frac{2}{\cancel{d}}\left(\frac{2+\cancel{d}-2}{2+d}\right) = -\frac{2}{2+d}}} \cdot \left(\frac{d}{2}\right)^{\frac{2}{2+d}}$$

$$= n^{-\frac{2}{2+d}} \left\{ \left(\frac{d}{2}\right)^{-\frac{d}{2+d}} + \left(\frac{d}{2}\right)^{\frac{2}{2+d}} \right\}$$

$$= \left(\frac{1}{n}\right)^{\frac{2}{2+d}} \left\{ \left(\frac{d}{2}\right)^{-\frac{d}{2+d}} + \left(\frac{d}{2}\right)^{\frac{2}{2+d}} \right\}$$

$$= \left(\frac{1}{n}\right)^{\frac{2}{2+d}} \left\{ \left(\frac{d}{2}\right)^{-\frac{d}{2+d}} + \left(\frac{d}{2}\right)^{\frac{2}{2+d}} \right\}$$

$$= \left(\frac{1}{n}\right)^{\frac{2}{2+d}} \cdot \left(\frac{d}{2}\right)^{\frac{2}{2+d}} \left(1 + \left(\frac{d}{2}\right)^{-1}\right)$$

$$= \left(\frac{1}{n}\right)^{\frac{2}{2+d}} \cdot \left(\frac{d}{2}\right)^{\frac{2}{2+d}} \left(\frac{2+d}{d}\right)$$

7) $$\Delta R = \left(\frac{1}{n}\right)^{\frac{2}{2+d}} \cdot \frac{d+2}{d} \left(\frac{d}{2}\right)^{\frac{2}{d+2}}$$

$$\Leftrightarrow \quad n^{\frac{2}{2+d}} = \frac{d+2}{d\Delta R} \left(\frac{d}{2}\right)^{\frac{2}{d+2}}$$

$$\Leftrightarrow \quad n = \left(\frac{d+2}{d\Delta R} \left(\frac{d}{2}\right)^{\frac{2}{d+2}}\right)^{\frac{2+d}{2}} = \left(\frac{d}{2}\right) \cdot \left(\frac{d+2}{d\Delta R}\right)^{\frac{2+d}{2}}$$