سوال ۱، الف) ویژگی ↑ ← pothesis space ↑

* پیچیدگی مدل یا توابع ↑ ← hypothesis space ↑

* افزایش داده های ورودی

ب) با حذف ویژگی‌های همبسته، مدل می‌تواند الگوهای پیچیده‌تر را بهتر تشخیص کند / fit بهتر و bias کمتر

" " " " " ممکن است توانایی کمتری برای تعمیم به بقیه داده‌ها داشته باشد (جدید)

که باعث var ↑ می‌شود.

ب ۲) ۱) در این حالت bias زیاد است و مدل نمی‌تواند ویژگی‌های داده را پیدا کند و افزایش داده

صرفاً باعث var ↓ می‌شود و bias را کاهش می‌دهد

۲) ممکن است باعث رسیدن به مدل بهتری شود.

سوال ۲) الف) خطا زمانی اتفاق می‌افتد که چند data با fهای یکسان labelهای متفاوت بدهند.

سه داده {$A=0, B=1, C=1$} ⟵ 2 label : 1 / 1 label : 0 ⟶ پیش‌بینی است یک خواهد بود و یک خطا اتفاق می‌افتد.

دو داده ($A=1, B=1, C=1$) ⟵ 1 label : 1 / 1 label : 1 ⟶ پیش‌بینی درست به صورت رندوم و یک خطا اتفاق می‌افتد.

پس برای B : $۲/۱۳ \times ۱۰۰ \simeq ۱۵٫۳۸ \%$

ب) وقتی f جداکننده‌ای نداریم ⟶ حداکثر خطا

تعداد کلاس‌ها و داده را k و c فرض کنید و وقتی مقدار fهای همه داده‌ها یکسان است حداکثر تعداد درست $\lceil \frac{c}{k} \rceil$ است و پس خطای $\frac{c - \lceil c/k \rceil}{c}$ را خواهیم داشت

CamScanner

سوال ۳) الف) K کلاس به صورت $w_1 \text{———} w_K$

$X : R^n$ و $X \to Y$ به طوری که R^n تو w_i ها داشته شده

$$P(Y=K \mid X=x) = \frac{e^{w_K^T x}}{\sum_{i=1}^{K} e^{w_i^T x}}$$

$K = 1, 2, \text{———}, K$

$w_K = 0$

$$P(Y=K \mid X=x) = \frac{e^{w_K^T x}}{1+\sum_{i=1}^{K-1} e^{w_i^T x}} \quad , \quad K = 1, \text{———}, K$$

برای $K=K$ چون $w_K = 0$ $\quad P(Y=K \mid X=x) = \dfrac{1}{1+\sum_{i=1}^{K-1} e^{w_i^T x}}$ برقرار است

ب) باید $\{w_1 \text{———} w_{K-1}\}$ تخمین زده شود

ب ۱)

$$l(w_1 \text{———} w_{K-1}) = \sum_{i=1}^{n} \ln P(Y=y_i \mid X=x_i)$$

$$= \sum_{i=1}^{n} \ln \frac{e^{w_{y_i}^T x_i}}{1+\sum_{l=1}^{K-1} e^{w_l^T x_i}} = \sum_{i=1}^{n}\left(w_{y_i}^T x_i - \ln\left(1+\sum_{l=1}^{K-1} e^{w_l^T x_i}\right)\right)$$

ب ۲)

$$\nabla l(w_K) = \sum_{i=1}^{n}\left(I(y_i=K)\, x_i - \frac{e^{w_K^T x_i}}{1+\sum_{l=1}^{K-1} e^{w_l^T x_i}} \right)$$

$$= \sum_{i=1}^{n}\left(I(y_i=K) - P(Y=K \mid X=x_i)\right) x_i$$

if $y_i = K \longrightarrow I(y_i=K) = 1$, $O.w. \rightsquigarrow 0$

ج)

$$\nabla f(w_k) = \nabla \ell(w_k) - \frac{\lambda}{2}(2w_k)$$

$$\longrightarrow \nabla f(w_k) = \nabla \ell(w_k) - \lambda w_k$$

CamScanner

سوال ۴

الف) در رابطه رگرسیون خطی

$$w_j = (x_j^T x_j)^{-1} x_j y \rightarrow w_j = \frac{x_j^T y}{x_j^T x_j}$$

در واقع $l = \frac{\|y - Xw\|^2}{2}$ باید مینیمم شود

$$l = \frac{1}{2}\left(y^T y - y^T X w - w^T X^T y + w^T X^T X w \right)$$

$$\xrightarrow{\nabla l = 0} \frac{1}{2}\left(-2X^T y + 2X^T X w \right) = 0 \longrightarrow w = (X^T X)^{-1} X^T y$$

ب) ستون های متعامد ← det صفر

$$X^T X = \mathrm{diag}\left((x_1 \cdot x_1), \ldots, (x_m \cdot x_m)\right)$$

$$\longrightarrow (X^T X)^{-1} = \mathrm{diag}\left((x_1^T x_1)^{-1}, \ldots, (x_m^T x_m)^{-1}\right)$$

$$w = (X^T X)^{-1} X^T y = \mathrm{diag}\left((x_1^T x_1)^{-1} \ldots \right) X^T y$$

$$\longrightarrow w_j = \left(\mathrm{diag}\left((x_1^T x_1)^{-1}, \ldots\right)\right)_j (X^T y)_j$$

$$= (w_j^T w_j)^{-1} (X^T y)_j = \frac{(X^T)_j y}{x_j^T x_j}$$

$$\longrightarrow w_j = \frac{x_j^T y}{x_j^T x_j}$$

اثبات شد!

پ)

$$X = \begin{bmatrix} x_j & \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \end{bmatrix} \quad X^T X = \begin{pmatrix} x_j^T x_j & \mathrm{sum}(x_j) \\ \mathrm{sum}(x_j) & n \end{pmatrix}$$

$$\longrightarrow (X^T X)^{-1} = \frac{1}{n\|x_j\|^2 - \mathrm{sum}^2(x_j)} \begin{pmatrix} n & -\mathrm{sum}(x_j) \\ -\mathrm{sum}(x_j) & x_j^T x_j \end{pmatrix}$$

$$X^T y = \begin{bmatrix} x_j^T y \\ \mathrm{sum}(y) \end{bmatrix}$$

$$(w_j, w_0) = w = (X^TX)^{-1}X^Ty = \frac{1}{n\|x_j\|^2 - sum^2(x_j)} \begin{pmatrix} n & -sum(x_j) \\ -sum(x_j) & x_j^T x_j \end{pmatrix} \begin{pmatrix} x_j^T y \\ sum(y) \end{pmatrix}$$

$$\longrightarrow w_j = \frac{\frac{x_j^T y}{n} - \frac{sum(x_j)}{n} \times \frac{sum(y)}{n}}{\frac{\|x_j\|^2}{n} - \left(\frac{sum(x_j)}{n}\right)^2} = \frac{E(x_j y) - E(x_j)E(y)}{E(x_j^2) - E^2(x_j)}$$

$$= \frac{Cov(x_j, y)}{Var(x_j)}$$

$$w_0 = \frac{E[y]E[x_j^2] - E[x_j]E[x_j y] - E[y]E[x_j]^2 + E[y]E[x_j]^2}{E[x_j^2] - E(x_j)^2}$$

$$= \frac{E(y)\,Var(x_j)}{Var(x_j)} + \frac{E(x_j)\left(E(y)E(x_j) - E(x_j y)\right)}{Var(x_j)}$$

$$= E(y) + E(x_j) \times \frac{-Cov(x_j, y)}{Var(x_j)}$$

$$\longrightarrow w_0 = E(y) - w_j E(x_j)$$

CamScanner