Nome: Luís Felipe de Melo Costa Silva
Número USP: 9297961

Lista de Exercícios 2 - MAC0460

Exercício 4

Queremos que $\epsilon(M, N, \delta) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} \leq 0.05$. Teremos:

$$\begin{aligned}
\sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} &\leq 0.05 \\
\frac{1}{2N} \ln \frac{2M}{\delta} &\leq 0.05^2 \\
\frac{1}{2N} &\leq \frac{0.05^2}{\ln \frac{2M}{\delta}} \\
2N &\leq \frac{\ln \frac{2M}{\delta}}{0.05^2} \\
N &\leq \frac{\ln \frac{2M}{\delta}}{2 \cdot 0.05^2}
\end{aligned}$$

Como $\delta = 0.03$:

a) Para $M = 1, N \leq \frac{\ln \frac{2}{0.03}}{2 \cdot 0.05^2} \cong 840$

b) Para $M = 100, N \leq \frac{\ln \frac{200}{0.03}}{2 \cdot 0.05^2} \cong 1761$

c) Para $M = 10000, N \leq \frac{\ln \frac{20000}{0.03}}{2 \cdot 0.05^2} \cong 2683$

Exercício 7

Temos que $y(x) = f(x) + \epsilon$, onde ϵ é uma v.a. com $\mathbb{E}(\epsilon) = 0$ e var$(\epsilon) = \mathbb{E}[\epsilon^2] = \sigma^2$. Para auxiliar nos cálculos, vamos definir uma hipótese média $\bar{g}(x) = \mathbb{E}_{\mathcal{D}}[g^{\mathcal{D}}(x)]$.

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}}[E_{out}(g^{(\mathcal{D})})] &= \mathbb{E}_{\mathcal{D}}[\mathbb{E}_{X,Y}[(g^{(\mathcal{D})}(x) - y(x))^2]] \\
&= \mathbb{E}_{X,Y}[\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - y(x))^2]] \\
&= \mathbb{E}_{X,Y}[\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x) + \bar{g}(x) - y(x))^2]] \\
&= \mathbb{E}_{X,Y}[\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x))^2 + (\bar{g}(x) - y(x))^2 + 2 \cdot \overbrace{(g^{(\mathcal{D})}(x) - \bar{g}(x))}^{0} \cdot (\bar{g}(x) - y(x))]] \\
&= \mathbb{E}_{X,Y}[\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x))^2 + (\bar{g}(x) - y(x))^2]] \\
&= \mathbb{E}_{X,Y}[\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x))^2] + \mathbb{E}_{\mathcal{D}}[(\bar{g}(x) - y(x))^2]] \\
&= \mathbb{E}_{X,Y}[\mathbb{E}_{\mathcal{D}}[\text{var}(x) + (\bar{g}(x) - y(x))^2]] \\
&= \text{var} + \mathbb{E}_{X,Y}[(\bar{g}(x) - y(x))^2] \\
&= \text{var} + \mathbb{E}_{X,Y}[(\bar{g}(x) - f(x) - \epsilon)^2] \\
&= \text{var} + \mathbb{E}_{X,Y}[(\bar{g}(x))^2 + (f(x))^2 + \epsilon^2 - 2 \cdot \bar{g}(x) \cdot f(x) - 2 \cdot \bar{g}(x) \cdot \epsilon + 2 \cdot f(x) \cdot \epsilon] \\
&= \text{var} + \mathbb{E}_{X,Y}[(\bar{g}(x) - f(x))^2 + \epsilon^2 - 2 \cdot \bar{g}(x) \cdot \epsilon + 2 \cdot f(x) \cdot \epsilon] \\
&= \text{var} + \mathbb{E}_{X,Y}[\text{bias}(x)] + \mathbb{E}_{X,Y}[\epsilon^2] - 2 \cdot \mathbb{E}_{X,Y}[\bar{g}(x) \cdot \epsilon] + 2 \cdot \mathbb{E}_{X,Y}[f(x) \cdot \epsilon] \\
&\overset{ind.}{=} \text{var} + \text{bias } + \mathbb{E}[\epsilon^2] - 2 \cdot \mathbb{E}_X[\bar{g}(x)] \cdot \overbrace{\mathbb{E}[\epsilon]}^{0} + 2 \cdot \mathbb{E}_X[f(x)] \cdot \overbrace{\mathbb{E}[\epsilon]}^{0} = \sigma^2 + \text{bias} + \text{var}
\end{aligned}
$$

$\square$

Exercício 8

a) Do enunciado, temos que o conjunto $\mathcal{H}$ possui funções no estilo $ax+b$. Logo, usando a definição de $\bar{g}(x)$, e definindo como K o número total de *data sets*:

$$
\begin{aligned}
\bar{g}(x) &= \mathbb{E}_{\mathcal{D}}[g^{(\mathcal{D})}(x)] = \mathbb{E}_{\mathcal{D}}[a^{(\mathcal{D})} \cdot x + b^{(\mathcal{D})}] = \mathbb{E}_{\mathcal{D}}[a^{(\mathcal{D})} \cdot x] + \mathbb{E}_{\mathcal{D}}[b^{(\mathcal{D})}] \\
&= \frac{1}{K}\sum_{k=1}^{K} a_k \cdot x + \frac{1}{K}\sum_{k=1}^{K} b_k = \frac{1}{K}\sum_{k=1}^{K} a_k \cdot x + b_k
\end{aligned}
$$

b) Fixando um número K de iterações, poderíamos fazer, a cada iteração:

- Gerar dois valores, x_1 e x_2, cada um retirado de uma distribuição Uniforme$(-1, 1)$. Com isso, geramos dois pontos, usando a $f(x)$, e temos nosso *data set*.
- Rodar o algoritmo de aprendizado, que resultará em uma $g(x)$.

Teremos K *data sets* e K funções $g(x)$. Com isso, podemos calcular:

- $\bar{g}(x)$, com a média das $g(x)$.
- O *bias*, com o valor $\mathbb{E}_X[(\bar{g}(x) - f(x))^2]$.
- A var, fazendo $\mathbb{E}_X[\mathbb{E}_{(\mathcal{D})}(g^{\mathcal{D}}(x) - \bar{g}(x))^2]]$
- Por último, E_{out}, com a fórmula $\mathbb{E}_{\mathcal{D}}[E_{out}(g^{(\mathcal{D})})] = \mathbb{E}_{\mathcal{D}}[\mathbb{E}_X[(g^{(\mathcal{D})}(x) - f(x))^2]]$

Exercício 10

a) Vamos chamar de $h(x)$ a saída do nosso algoritmo e C a variável que representa o custo. Teremos:

- cost(accept) = $\mathbb{E}(C|h(x) = 1) = g(x) \cdot 0 + (1 - g(x)) \cdot c_a = (1 - g(x)) \cdot c_a$
- cost(reject) = $\mathbb{E}(C|h(x) = -1) = g(x) \cdot c_r + (1 - g(x)) \cdot 0 = g(x) \cdot c_r$

b) A regressão logística gera uma saída num intervalo contínuo $[0, 1]$. Quando usamos esse resultado para uma classificação binária, precisamos de um limiar κ. O custo de aceite é igual ao custo de rejeite nesse ponto, já que o intervalo é contínuo. Então, fazendo $g(x) = \kappa$, teremos:

$$
\begin{gathered}
\text{cost(reject) = cost(accept)} \\
(1 - g(x)) \cdot c_a = g(x) \cdot c_r \\
(1 - \kappa) \cdot c_a = \kappa \cdot c_r \\
c_a - \kappa \cdot c_a = \kappa \cdot c_r \\
\kappa = \frac{c_a}{c_a + c_r}
\end{gathered}
$$

$\square$

c) Para o exemplo do supermercado, temos que $c_a = 1$ e $c_r = 10$. Portanto, teremos $\kappa = \frac{1}{11}$, que indica um bom intervalo de aceite, já que é o falso negativo que incomoda. Já para a CIA, com $c_a = 1000$ e $c_r = 1$, $\kappa = \frac{1000}{1001}$, e nos dá um intervalo bem restritivo para o aceite, de modo a evitar ao máximo falsos positivos.