Subject: Date:

[illegible] ۴۰۲۰۷۱۸۴ [illegible]

[illegible] VC dimension [illegible] shattering [illegible]

[illegible] H [illegible] C [illegible] X [illegible] shatter [illegible]

[illegible] H [illegible] C [illegible] H_C [illegible]

[illegible] $H_C = \{(h(c_1), h(c_2), \ldots, h(c_m)) \mid h \in H\}$

VC dimension [illegible] H [illegible] C [illegible]

[illegible] X [illegible] H shatter [illegible]

[illegible] VC dimension [illegible] H [illegible] d [illegible]

۱) [illegible] C [illegible] X [illegible] d [illegible] H shatter [illegible]

۲) [illegible] C [illegible] X [illegible] $d+1$ [illegible] H shatter [illegible]

[illegible] X [illegible]

[illegible] sample complexity [illegible]

الف) $\text{sign}(x^T x, \theta) = \text{sign}(\|x\|_2^2, \theta)$

VC dimension [illegible]

۱) [illegible] shatter [illegible]

[illegible] $\theta < \|x\|_2^2$ [illegible] $\theta > \|x\|_2^2$ [illegible]

۲) [illegible] shatter [illegible]

[illegible] $\|x\|_2 = \|y\|_2$ [illegible]

[illegible] $\|x\|_2 < \|y\|_2$ [illegible]

۲) $\text{sign}(x_1\theta_1 + x_2\theta_2 + \theta_0)$

VC dimension [illegible] ۳ [illegible]

۱) [illegible] shatter [illegible]

۲) [illegible] shatter [illegible]

[illegible]

[illegible]

Subject: Date:

حالت پیچیده‌تر ... نشان ... است ... مشخص شده ... است ... حالت

پیچیده‌تر ... نشان ... را نشان دادیم که ...

ع) ابتدا ثابت می‌کنیم VC dimension برای پرسپترون بدون بایاس در فضای N بعدی برابر N است.

۱) کافیست نقاط $e_1, e_2, \dots, e_N$ را در نظر بگیریم. این نقاط shatter می‌شوند زیرا برای هر

برچسب‌گذاری $y_1, y_2, \dots, y_N$ اگر قرار دهیم $w_i = y_i$ خواهیم داشت $\forall i \in \{1, \dots, N\}$, $\langle w, e_i \rangle = y_i$

۲) فرض کنید $N+1$ نقطه $x_1, x_2, \dots, x_{N+1}$ داریم که shatter می‌شوند. چون $N+1$ بردار در فضای N بعدی

مستقل خطی نیستند، ضرایب $a_1, a_2, \dots, a_{N+1}$ که همه صفر نیستند وجود دارند به طوری که $\sum_{i=1}^{N+1} a_i x_i = 0$

تعریف می‌کنیم $I = \{i : a_i > 0\}$, $J = \{j : a_j < 0\}$

مثلاً فرض کنید I ناتهی باشد. آنگاه داریم

$\sum_{i \in I} a_i x_i = \sum_{j \in J} |a_j| x_j$

حال برچسب‌گذاری را در نظر بگیرید که shatter شدن w را نشان می‌دهد

$\langle w, x_i \rangle > 0 \quad \forall i \in I$, $\langle w, x_j \rangle < 0 \quad \forall j \in J$

در این صورت داریم

$0 < \langle \sum_{i \in I} a_i x_i, w \rangle = \langle \sum_{i \in I} a_i x_i, w \rangle = \langle \sum_{j \in J} |a_j| x_j, w \rangle = \sum_{j \in J} |a_j| \langle x_j, w \rangle < 0$

... که تناقض است. ... اگر I تهی باشد ... J ... به همین ترتیب ...

... تناقض است.

حال ثابت می‌کنیم VC dimension برای پرسپترون با بایاس در فضای N بعدی برابر $N+1$ است.

۱) کافیست نقاط $0, e_1, \dots, e_N$ را در نظر بگیریم. این نقاط shatter می‌شوند زیرا برای هر برچسب‌گذاری $y_1, y_2, \dots$

y_N, y_{N+1} اگر قرار دهیم $w_i = y_i \; \forall i \in \{1, \dots, N\}$, $w_{N+1} = \frac{1}{N} y_{N+1}$ خواهیم داشت

$\langle w, e_i \rangle + w_{N+1} = y_i + \frac{1}{N} y_{N+1} \quad \forall i \in \{1, \dots, N\}$, $\langle w, 0 \rangle + w_{N+1} = \frac{1}{N} y_{N+1}$

۲) فرض کنید $N+2$ نقطه $x_1, x_2, \dots, x_{N+2}$ داریم که shatter می‌شوند. در این صورت ...

... این نقاط ... با ... بعد ... بردارهای ... N+1 بعدی ... shatter ...

... shatter ... که ... است ... تناقض ...

... VC dimension پرسپترون با بایاس در فضای N بعدی برابر $N+1$ است.

Subject: Date:

سوال ۲) [illegible] MSE [illegible] Loss Function [illegible]

$$w^* = \arg\min_w \sum_{i=1}^{m} (y_i - x_i^T w)^2$$

[illegible] ML [illegible]

[illegible]: $y_i = x_i^T w + \varepsilon_i \;,\; \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

[illegible]

$$w^* = \arg\max_w \prod_{i=1}^{m} f(y_i | x_i, w, \sigma^2) = \arg\max_w \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - x_i^T w)^2}{2\sigma^2}} =$$

$$\arg\max_w \log\left(\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - x_i^T w)}{2\sigma^2}}\right) = \arg\max_w \sum_{i=1}^{m} \left(-\log(\sqrt{2\pi}\sigma) - \frac{(y_i - x_i^T w)^2}{2\sigma^2}\right) =$$

$$\arg\max_w \sum_{i=1}^{m} \left(-\frac{(y_i - x_i^T w)^2}{2\sigma^2}\right) = \arg\max_w \sum_{i=1}^{m} \left(-(y_i - x_i^T w)\right) = \arg\min_w \sum_{i=1}^{m} (y_i - x_i^T w)^2$$

[illegible]

سوال ۳) [illegible] μ [illegible] σ^2 [illegible] ML [illegible]

$$(\mu^*, \sigma^{2*}) = \arg\max_{\mu, \sigma^2} \prod_{i=1}^{m} f(x_i | \mu, \sigma^2) = \arg\max_{\mu, \sigma^2} \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} =$$

$$\arg\max_{\mu, \sigma^2} \sum_{i=1}^{m} \left(-\log(\sqrt{2\pi}\sigma) - \frac{(x_i - \mu)^2}{2\sigma^2}\right) \longrightarrow$$

$$\frac{\partial}{\partial \mu} = 0 \longrightarrow \frac{1}{2\sigma^2} \sum_{i=1}^{m} 2(x_i - \mu) = 0 \longrightarrow \mu = \frac{1}{m} \sum_{i=1}^{m} x_i$$

$$\frac{\partial}{\partial \sigma^2} = 0 \longrightarrow \sum_{i=1}^{m} \left(-\frac{1}{2\sigma^2} + \frac{(x_i - \mu)^2}{2\sigma^4}\right) = 0 \longrightarrow \sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu)^2$$

سوال ۴) [illegible] Regression [illegible]

$$w^* = \arg\min_w \|Y - XW\|_2^2$$

[illegible] X [illegible] $X^T X$ [illegible]

$$w^* = \arg\min (Y - XW)^T (Y - XW) = \arg\min_w Y^T Y - Y^T X W - W^T X^T Y + W^T X^T X W$$

$$\frac{\partial}{\partial W} = 0 \longrightarrow -X^T Y - X^T Y + 2 X^T X W = 0 \longrightarrow X^T X W = X^T Y \longrightarrow W = (X^T X)^{-1} X^T Y$$

[illegible] Regularization [illegible]

$$w^* = \arg\min_w \|Y - XW\|_2^2 + \lambda \|W\|_2^2 =$$

$$\arg\min_w Y^T Y - Y^T X W - W^T X^T Y + W^T X^T X W + \lambda W^T W$$

$$\frac{\partial}{\partial W} = 0 \longrightarrow -X^T Y - X^T Y + 2 X^T X W + 2\lambda I W = 0 \longrightarrow (X^T X + \lambda I) W = X^T Y \longrightarrow$$

$$W = (X^T X + \lambda I)^{-1} X^T Y$$