CS 4780/5780: Introduction to Machine Learning **(Due: 09/29/20)**
Copyrighted material, do not share.

Assignment #2: Hypothesis Testing, Linear Classifiers, Perceptrons

Instructor: Nika Haghtalab, Thorsten Joachims *Name:* Kevin Klaben, *Netid:* kek228

Course Policy: Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- Please include your name and NetIDs on the first page. We recommend typesetting your submission in LaTeX, and an Overleaf template is linked on the Canvas module. When submitting, remember to mark which page has which response.
- Assignments are due by 5pm on the due date in PDF form on Gradescope.
- Late assignments can be submitted on Gradescope up to Sunday, Oct 4 at 5pm. This is also when the solutions will be released.
- You can do this assignment in groups of 1-2. Please submit no more than one submission per group. Collaboration across groups is not permitted.
- All sources of material outside the course must be cited. The University Academic Code of Conduct will be strictly enforced.

Submission Instructions: All group members must be added to the Gradescope submission. If you're the one submitting, add your group members on Gradescope. Otherwise, make sure you are added to the submission. Put your names on the PDF, this helps us track the groups in case there are errors on Gradescope.

Problem 1

(a)i

1(a)i)

$$P(X_i) = \begin{cases} p & X_i = 1 \\ 1-p & X_i = 0 \end{cases} \qquad p_s = \frac{n_1}{n} \; \frac{139}{285} = 0.4877$$

X_i corresponds to the doctor's diagnosis of patient i, the test data.

$\bar{X}$ is the sample probability of X_i being equal to 1.

n is the number of images evaluated by the doctor while n_1 is the number in this sample classified by the doctor as 1.

(a)ii

ii.) $\alpha = P(|\bar{X} - \mathbb{E}[\bar{X}]| > \epsilon) \leq 2\exp(-2n\epsilon^2)$

$$\alpha \leq 2\exp(-2n\epsilon^2)$$

$$\frac{\alpha}{2} \leq e^{-2n\epsilon^2}$$

$$\log\left(\frac{\alpha}{2}\right) \leq -2n\epsilon^2$$

$$-\frac{\log\left(\frac{\alpha}{2}\right)}{2n} \leq \epsilon^2$$

$$\sqrt{-\frac{\log\left(\frac{\alpha}{2}\right)}{2n}} \leq \epsilon$$

ϵ is deviation from mean thus interval is +/- this amount from sample mean

$$err_P(\hat{h}) \in \left[err_S(\hat{h}) - \sqrt{\frac{-0.5\log\left(\frac{\alpha}{2}\right)}{n}},\ err_S(\hat{h}) + \sqrt{\frac{-0.5\log\left(\frac{\alpha}{2}\right)}{n}}\right]$$

(a)iii

iii $err_p(h_A) = \left[err_s(h_A) - \sqrt{\frac{-0.5\log(\frac{\alpha}{2})}{n}}, err_s(h_A) \sqrt{\frac{-0.5\log(\frac{\alpha}{2})}{n}} \right]$

$err_p(h_A) = \left[\frac{30}{285} - \sqrt{\frac{-0.5\log(\frac{0.025}{2})}{285}}, \frac{30}{285} + \sqrt{\frac{-0.5\log(\frac{0.025}{2})}{285}} \right]$

$err_p(h_A) = \left[0.0478, 0.16305 \right]$ ← 97.5% CI for h_A

$err_p(h_B) = \left[err_s(h_B) - \sqrt{\frac{-0.5\log(\frac{\alpha}{2})}{n}}, err_s(h_B) + \sqrt{\frac{-0.5\log(\frac{\alpha}{2})}{n}} \right]$

$= \left[\frac{36}{285} - \sqrt{\frac{-0.5\log(\frac{0.025}{2})}{285}}, \frac{36}{285} + \sqrt{\frac{-0.5\log(\frac{0.025}{2})}{285}} \right]$

$= \left[0.0853, 0.18410 \right]$ ← 97.5% CI for h_B

(b)

(b) $H_0: err_p(\hat{h}_A) = err_p(\hat{h}_B)$

$H_A: err_p(\hat{h}_A) \neq err_p(\hat{h}_B)$

wins h_A over h_B: 15

wins h_B over h_A: 9

$P(W \leq w \mid p = 0.5, n = w+1)$

$P(W \leq 9 \mid p = 0.5, n = 24)$

$= \sum_{i=0}^{i=9} \binom{24}{i} (0.5)^{24} = 2^9 (0.5)^{24} = 3.0518 \times 10^{-5}$

as $p = 3.0518 \times 10^{-5} < 0.1$, we reject the null hypothesis and conclude that the true error rates of h_A and h_B are significantly different

Problem 2

(a)

2.) (a) $h_{w,b}(x) = \begin{cases} +1, & \text{if } w \cdot x + b > 0 \\ -1, & \text{otherwise} \end{cases}$

$w = -1 \quad b = 0$

(b)

(b)

$h_{\vec{w},b}(\vec{x}) = \begin{cases} +1, & \text{if } \vec{w} \cdot \vec{x} + b > 0 \\ -1, & \text{otherwise} \end{cases}$

$\vec{w} = (1, \ldots, 1)$ length d

$b = -m$

(c)

(c) $\vec{w} = (\underbrace{1, \ldots, 1}_{1-10=1}, \underbrace{0, \ldots, 0}_{11-90=0}, \underbrace{-1, \ldots, -1}_{91-100=-1})$

$b = -5$

(d)

(d) $\vec{w} = (1, 1)$

$b = 1.5$

	0	1
0	−1	−1
1	−1	1

(e)

(e) $\vec{w} = (c_1 2^{t}, c_2 2^{t-1}, \ldots, c_{t-1} 2^{t-(t-2)}, c_t 2^{t-(t-1)}, 0, \ldots 0) \times (d-t+1)$

$b = -(d-t+1)$

Problem 3

(a)

3 (a) $\vec{w}^{(0)} = (0,0)$

while there is $i \in [1,2,3]$ s.t. $y_i(\vec{w}^{(t)} \cdot \vec{x}_i) \le 0$

$1(0 \cdot 1 + 0 \cdot 3) \le 0$

$y_1(w^{(0)} \cdot \vec{x}_1) \le 0$

$w^{(1)} = \vec{w}^{(0)} + y_1 x_1$

$= (0,0) + 1(1,3)$

$\vec{w}^{(1)} = (1,3)$

$y_1(w^{(1)} \cdot \vec{x}_1) = 1(1 \cdot 1 + 3 \cdot 3) = 10 \not\le 0$

$y_2(w^{1} \cdot \vec{x}_2) = -1(1 \cdot -1 + 3 \cdot 3) = -8 \le 0$

$w^{(2)} = w^{(1)} + y_2 \vec{x}_2 = (1,3) + -1(-1,3) = (2,0)$

$y_1(w^{(2)} \cdot \vec{x}_1) = 1(2 \cdot 1 + 0 \cdot 3) = 2 \not\le 0$

$y_2(w^{(2)} \cdot \vec{x}_2) = -1(2 \cdot -1 + 0 \cdot 3) = 2 \not\le 0$

$y_3(w^{(2)} \cdot \vec{x}_3) = 1(2 \cdot 0 + 0 \cdot 1) = 0 \le 0$

$w^{(3)} = w^{(2)} + y_3 \vec{x}_3 = (2,0) + 1(0,1) = (2,1)$

$y_1(w^{(3)} \cdot \vec{x}_1) = 1(2 \cdot 1 + 1 \cdot 3) = 5 \not\le 0$

$y_2(w^{(3)} \cdot \vec{x}_2) = -1(2 \cdot -1 + 1 \cdot 3) = -1 \le 0$

$w^{(4)} = w^{(3)} + y_2 \vec{x}_2 = (2,1) + -1(-1,3) = (3,-2)$

$y_1(w^{(4)} \cdot \vec{x}_1) = 1(3 \cdot 1 + -2 \cdot 3) = -3 \le 0$

$w^{(5)} = w^{(4)} + y_i \vec{x}_i = (3,-2) + 1(1,3) = (4,1)$

$y_1(w^{(5)} \cdot \vec{x}_1) = 1(4 \cdot 1 + 1 \cdot 3) = 7 \not\le 0$

$y_2(w^{(5)} \cdot \vec{x}_2) = -1(4 \cdot -1 + 1 \cdot 3) = 1 \not\le 0$

$y_3(w^{(5)} \cdot \vec{x}_3) = 1(4 \cdot 0 + 1 \cdot 1) = 1 \not\le 0$

$w^{(0)} = (0,0)$

$w^{(1)} = (1,3)$

$w^{(2)} = (2,0)$

$w^{(3)} = (2,1)$

$w^{(4)} = (3,-2)$

$w^{(5)} = (4,1)$

(b)

(b) $w^{(t+1)} = w^{(t)} + y_i \vec{x}_i$, thus if we know the number of times each i is used to update, we can simply multiply the addition that many times and sum.

$$w^{(final)} = w^{(0)} + (1) y_1 \vec{x}_1 + (3) y_2 \vec{x}_2 + (1) y_3 \vec{x}_3 + (1) y_4 \vec{x}_4 + (1) y_5 \vec{x}_5$$

$$= (0,0,0,0,0) + (1)(1)(1,3,3,7,0) + 3(-1)(0,0,0,0,1)$$

$$+ 1(1)(0,0,0,1,0) + 1(1)(9,1,2,0,0) + 1(-1)(2,8,4,0,-2)$$

$$w^{(final)} = (8,-4,1,8,-1)$$