
Machine Learning

Answer Sheet for Homework 3

Da-Min HUANG

R04942045

Graduate Institute of Communication Engineering, National Taiwan University

November 1, 2015

Problem 1

Set $\sigma = 0.1$ and $d = 8$, then we can rewrite the formula to be

$$\mathbb{E}_{\mathcal{D}} [E_{\text{in}}(\mathbf{w}_{\text{lin}})] = 0.01 \left(1 - \frac{9}{N}\right) > 0.008 \Rightarrow 0.2 > \frac{9}{N} \Rightarrow N > 45 \quad (1)$$

□

Problem 2

- (a) \mathbf{H} is positive semi-definite \Leftrightarrow All eigenvalues is non-negative. Refer to choice (c), we have shown the properties.

- (b) Consider $\mathbf{X} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$, we have

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix} \quad (2)$$

which has no inverse.

Also, consider the physical meaning of \mathbf{H} . The inverse of projection is not injective, which implies that \mathbf{H} is not always invertible.

- (c) Refer to choice (e), we have $\mathbf{H}^2 = \mathbf{H}$. Suppose λ is the eigenvalue of some non-zero vector \vec{v} ,

$$\mathbf{H}\vec{v} = \lambda\vec{v} = \mathbf{H}^2\vec{v} = \mathbf{H}(\lambda\vec{v}) = \lambda(\mathbf{H}\vec{v}) = \lambda^2\vec{v} \Rightarrow \lambda^2 = \lambda \quad (3)$$

Hence, the possible results of λ is 1 or 0.

- (d) Consider the physical meaning of \mathbf{H} . Since there are d features, so at least $d + 1$ (since the feature vector contains x_0 term) eigenvalues are 1.

- (e) By the definition of \mathbf{H} , we have

$$\mathbf{H}^2 = \left(\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \left(\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \quad (4)$$

$$= \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \left((\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \right) \mathbf{X}^T \quad (5)$$

$$= \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \mathbf{H} \quad (6)$$

So

$$\mathbf{H}^2 = \mathbf{H} \Rightarrow \mathbf{H}^{1126} = \mathbf{H} \quad (7)$$

□

Problem 3

If $\text{sign}(\mathbf{w}^T \mathbf{x}) \neq y$, then $y\mathbf{w}^T \mathbf{x} < 0$ since the sign of y and $\mathbf{w}^T \mathbf{x}$ are different. Similarly, if $\text{sign}(\mathbf{w}^T \mathbf{x}) = y$, then $y\mathbf{w}^T \mathbf{x} \geq 0$.

Claim: $(\max(0, 1 - y\mathbf{w}^T \mathbf{x}))^2$ is an upper bound.

Proof of claim:

Consider the following cases.

$$1. [\text{sign}(\mathbf{w}^T \mathbf{x}) \neq y] = 0$$

Then $y\mathbf{w}^T \mathbf{x} \geq 0$. Hence $(\max(0, 1 - y\mathbf{w}^T \mathbf{x}))^2 \geq 0$, which bounds $[\text{sign}(\mathbf{w}^T \mathbf{x}) \neq y]$.

$$2. [\text{sign}(\mathbf{w}^T \mathbf{x}) \neq y] = 1$$

Then $y\mathbf{w}^T \mathbf{x} < 0$. Hence $(\max(0, 1 - y\mathbf{w}^T \mathbf{x}))^2 = (1 - y\mathbf{w}^T \mathbf{x})^2 > 1$, which bounds $[\text{sign}(\mathbf{w}^T \mathbf{x}) \neq y]$.

□

Problem 4

Set $y\mathbf{w}^T\mathbf{x} := z$. Consider $\max(0, -y\mathbf{w}^T\mathbf{x}) = \max(0, -z) := f(z)$. We have $f(z) = -z$ if $z \leq 0$, else $f(z) = 0$. So

$$\lim_{z \rightarrow 0^-} \frac{f(z) - f(0)}{z - 0} = \frac{-z - 0}{z - 0} = -1, \quad \lim_{z \rightarrow 0^+} \frac{f(z) - f(0)}{z - 0} = \frac{0 - 0}{z - 0} = 0 \quad (8)$$

Hence, $f(z)$ is not differentiable at $z = 0$.

□

Problem 5

Calim: $\max(0, -y\mathbf{w}^T\mathbf{x})$ results in PLA.

Proof of claim:

Consider following cases.

1. $y = \text{sign}(\mathbf{w}^T\mathbf{x})$. Then we have $y\mathbf{w}^T\mathbf{x} > 0$. So

$$\max(0, -y\mathbf{w}^T\mathbf{x}) = 0 \quad (9)$$

2. $y \neq \text{sign}(\mathbf{w}^T\mathbf{x})$. Then we have $y\mathbf{w}^T\mathbf{x} < 0$. So

$$\max(0, -y\mathbf{w}^T\mathbf{x}) = -y\mathbf{w}^T\mathbf{x} \Rightarrow -\nabla_{\mathbf{w}} \max(0, -y\mathbf{w}^T\mathbf{x}) = y\mathbf{x} \quad (10)$$

□

Problem 6

$$\nabla E(0, 0) = \left(\frac{\partial E}{\partial u}, \frac{\partial E}{\partial v} \right) \Big|_{(0,0)} \quad (11)$$

$$= (e^u + ve^{uv} + 2u - 2v - 3, 2e^{2v} + ue^{uv} - 2u + 4v - 2) \Big|_{(0,0)} \quad (12)$$

$$= (-2, 0) \quad (13)$$

□

Problem 7

$$(u_1, v_1) = (0, 0) - 0.01 \nabla E(0, 0) = (0.02, 0) \quad (14)$$

$$(u_2, v_2) = (0.02, 0) - 0.01 \nabla E(0.02, 0) \approx (0.039398, 0.0002) \quad (15)$$

$$(u_3, v_3) \approx (0.039398, 0.0002) - 0.01 \nabla E(0.039398, 0.0002) \quad (16)$$

$$\approx (0.0582102, 0.000577975) \quad (17)$$

$$(u_4, v_4) \approx (0.0764524, 0.00111381) \quad (18)$$

$$(u_5, v_5) \approx (0.09414, 0.00178911) \quad (19)$$

$$E(u_5, v_5) \approx 2.825 \quad (20)$$

□

Problem 8

$$\nabla E(0, 0) = \left(\frac{\partial E}{\partial u}, \frac{\partial E}{\partial v} \right) \quad (21)$$

$$= (e^u + ve^{uv} + 2u - 2v - 3, 2e^{2v} + ue^{uv} - 2u + 4v - 2) \quad (22)$$

From this we compute the Hessian matrix

$$\nabla^2 E(u, v) = \begin{pmatrix} e^u + v^2 e^{uv} + 2 & (uv + 1)e^{uv} - 2 \\ (uv + 1)e^{uv} - 2 & 4e^{2v} + u^2 e^{uv} + 4 \end{pmatrix} \quad (23)$$

So

$$\hat{E}(\Delta u, \Delta v) = E(0, 0) + \nabla E(0, 0) \cdot (\Delta u, \Delta v) + \frac{1}{2} (\Delta u, \Delta v) \nabla^2 E(0, 0) \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \quad (24)$$

$$= 3 - 2\Delta u + \frac{1}{2} (\Delta u, \Delta v) \begin{pmatrix} 3 & -1 \\ -1 & 8 \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \quad (25)$$

$$= \frac{3}{2} (\Delta u)^2 + 4 (\Delta v)^2 - \Delta u \Delta v - 2\Delta u + 0\Delta v + 3 \quad (26)$$

□

Problem 9

Claim: $-(\nabla^2 E(u, v))^{-1} \nabla E(u, v)$ is the Newton direction.

Proof of claim:

$$\frac{\partial \hat{E}(\Delta u, \Delta v)}{\partial (\Delta u, \Delta v)} = \nabla E(u, v) + \nabla^2 E(u, v) (\Delta u, \Delta v) = 0 \quad (27)$$

$$\Rightarrow (\Delta u, \Delta v) = -(\nabla^2 E(u, v))^{-1} \nabla E(u, v) \quad (28)$$

□

Problem 10

$$(u_1, v_1) \approx (0.695652173913, 0.0869565217391) \quad (29)$$

$$(u_2, v_2) \approx (0.613762221112, 0.0711078990173) \quad (30)$$

$$(u_3, v_3) \approx (0.611812859879, 0.0705000613365) \quad (31)$$

$$(u_4, v_4) \approx (0.611811717261, 0.0704995471019) \quad (32)$$

$$(u_5, v_5) \approx (0.61181171726, 0.0704995471016) \quad (33)$$

$$E(u_5, v_5) \approx 2.36082334564 \quad (34)$$

This equals to the value of Problem 7 after 746 updates.

□

Problem 11

Write a program Q11.py to test, $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6\}$ is the biggest subset that can be shattered by the union of quadratic, linear, or constant hypotheses of \mathbf{x} .

□

Problem 12

By the transform, we have $(\Phi(\mathbf{x}))_i = z_i = \left(0, \dots, \underbrace{1}_{i\text{-th term}}, \dots, 0\right)$. To shatter the original N points, we can assign w_i to be positive or negative to get $\mathbf{x}_i \circ$ or \times .

So, this transform shatter any N points. Hence $d_{vc}(\mathcal{H}_\Phi) = \infty$.

□

Problem 13

The average E_{in} is 0.503979.

□

Problem 14

The returned $\mathbf{w}_{\text{Lin}} = (-1.00134023, 0.075299620.01237623, 0.0812999, 1.69273348, 1.53664765)$.

□

Problem 15

The average $E_{\text{out}} = 0.127198$.

□

Problem 16

Sum the minimized negative log likelihood h_y , which is $\min_y (-\ln(h_y))$, we have

$$E_{\text{in}} = \frac{1}{N} \sum_{n=1}^N \left(-\ln \left(\frac{\exp(\mathbf{w}_{y_n}^T \mathbf{x}_n)}{\sum_{i=1}^K \exp(\mathbf{w}_i^T \mathbf{x}_n)} \right) \right) \quad (35)$$

$$= \frac{1}{N} \sum_{n=1}^N \left(\ln \left(\sum_{i=1}^K \exp(\mathbf{w}_i^T \mathbf{x}_n) \right) - \mathbf{w}_{y_n}^T \mathbf{x}_n \right) \quad (36)$$

□

Problem 17

$$\frac{\partial E_{\text{in}}}{\partial \mathbf{w}_i} = \frac{1}{N} \sum_{n=1}^N \frac{\partial}{\partial \mathbf{w}_i} \left(\ln \left(\sum_{i=1}^K \exp(\mathbf{w}_i^T \mathbf{x}_n) \right) - \mathbf{w}_{y_n}^T \mathbf{x}_n \right) \quad (37)$$

$$= \frac{1}{N} \sum_{n=1}^N \left(\frac{1}{\sum_{i=1}^K \exp(\mathbf{w}_i^T \mathbf{x}_n)} \frac{\partial}{\partial \mathbf{w}_i} \exp(\mathbf{w}_i^T \mathbf{x}_n) - \frac{\partial}{\partial \mathbf{w}_i} \mathbf{w}_{y_n}^T \mathbf{x}_n \right) \quad (38)$$

$$= \frac{1}{N} \sum_{n=1}^N \left(\frac{\exp(\mathbf{w}_i^T \mathbf{x}_n)}{\sum_{i=1}^K \exp(\mathbf{w}_i^T \mathbf{x}_n)} \mathbf{x}_n - \mathbb{I}[y_n = i] \mathbf{x}_n \right) \quad (39)$$

$$= \frac{1}{N} \sum_{n=1}^N (h_i(\mathbf{x}_n) - \mathbb{I}[y_n = i]) \mathbf{x}_n \quad (40)$$

□

Problem 18

The $E_{\text{out}} = 0.475$.

□

Problem 19

The $E_{\text{out}} = 0.220$.

□

Problem 20

The $E_{\text{out}} = 0.473$.

□

Reference

- [1] Lecture Notes by Hsuan-Tien LIN, Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan.