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

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

$\square$

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

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

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 $\vec{v}$,

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

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

$$\begin{aligned}
\mathbf{H}^2 &= \left(\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\left(\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right) && (4)\\
&= \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\left(\mathbf{X}^T\mathbf{X}\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)\mathbf{X}^T && (5)\\
&= \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T = \mathbf{H} && (6)
\end{aligned}$$

So

$$\mathbf{H}^2 = \mathbf{H} \Rightarrow \mathbf{H}^{1126} = \mathbf{H} \tag{7}$$

□

Problem 3

If $\text{sign}\left(\mathbf{w}^T\mathbf{x}\right) \neq y$, then $y\mathbf{w}^T\mathbf{x} < 0$ since the sign of y and $\mathbf{w}^T\mathbf{x}$ are different. Similarly, if $\text{sign}\left(\mathbf{w}^T\mathbf{x}\right) = y$, then $y\mathbf{w}^T\mathbf{x} \geq 0$.
Claim: $\left(\max\left(0, 1 - y\mathbf{w}^T\mathbf{x}\right)\right)^2$ is an upper bound.
Proof of claim:
Consider the following cases.

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

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

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

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

□

Problem 4

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

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

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

□

Problem 5

<u>Calim</u>: $\max\left(0, -y\mathbf{w}^T\mathbf{x}\right)$ results in PLA.
<u>Proof of claim</u>:
Consider following cases.

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

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

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

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

□

Problem 6

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

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

$$= (-2, 0) \tag{13}$$

□

Problem 7

$$(u_1, v_1) = (0,0) - 0.01\nabla E\,(0,0) = (0.02, 0) \tag{14}$$
$$(u_2, v_2) = (0.02, 0) - 0.01\nabla E\,(0.02, 0) \approx (0.039398, 0.0002) \tag{15}$$
$$(u_3, v_3) \approx (0.039398, 0.0002) - 0.01\nabla E\,(0.039398, 0.0002) \tag{16}$$
$$\approx (0.0582102, 0.000577975) \tag{17}$$
$$(u_4, v_4) \approx (0.0764524, 0.00111381) \tag{18}$$
$$(u_5, v_5) \approx (0.09414, 0.00178911) \tag{19}$$
$$E\,(u_5, v_5) \approx 2.825 \tag{20}$$

□

Problem 8

$$\nabla E(0,0) = \left(\frac{\partial E}{\partial u}, \frac{\partial E}{\partial v}\right) \tag{21}$$
$$= \left(e^u + ve^{uv} + 2u - 2v - 3, 2e^{2v} + ue^{uv} - 2u + 4v - 2\right) \tag{22}$$

From this we compute the Hessian matrix

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

So

$$\hat{E}\,(\Delta u, \Delta v) = E\,(0,0) + \nabla E\,(0,0) \cdot (\Delta u, \Delta v) + \frac{1}{2}\,(\Delta u, \Delta v)\,\nabla^2 E\,(0,0) \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \tag{24}$$
$$= 3 - 2\Delta u + \frac{1}{2}\,(\Delta u, \Delta v) \begin{pmatrix} 3 & -1 \\ -1 & 8 \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \tag{25}$$
$$= \frac{3}{2}\,(\Delta u)^2 + 4\,(\Delta v)^2 - \Delta u \Delta v - 2\Delta u + 0\Delta v + 3 \tag{26}$$

□

Problem 9

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

$$\frac{\partial \hat{E}\,(\Delta u, \Delta v)}{\partial\,(\Delta u, \Delta v)} = \nabla E\,(u,v) + \nabla^2 E\,(u,v)\,(\Delta u, \Delta v) = 0 \tag{27}$$
$$\Rightarrow (\Delta u, \Delta v) = -\left(\nabla^2 E\,(u,v)\right)^{-1} \nabla E\,(u,v) \tag{28}$$

□

Problem 10

$$(u_1, v_1) \approx (0.695652173913, 0.0869565217391) \tag{29}$$
$$(u_2, v_2) \approx (0.613762221112, 0.0711078990173) \tag{30}$$
$$(u_3, v_3) \approx (0.611812859879, 0.0705000613365) \tag{31}$$
$$(u_4, v_4) \approx (0.611811717261, 0.0704995471019) \tag{32}$$
$$(u_5, v_5) \approx (0.61181171726, 0.0704995471016) \tag{33}$$
$$E(u_5, v_5) \approx 2.36082334564 \tag{34}$$

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 $(\Phi(\mathbf{x}))_i = z_i = \left(0, \ldots, \underbrace{1}_{i\text{-th term}}, \ldots, 0\right)$. 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 $\min_y\left(-\ln\left(h_y\right)\right)$, we have

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

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

□

Problem 17

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

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

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

$$= \frac{1}{N}\sum_{n=1}^{N}\left(h_i\left(\mathbf{x}_n\right) - [\![y_n = i]\!]\right)\mathbf{x}_n \tag{40}$$

□

Problem 18

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

$\square$

Problem 19

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

$\square$

Problem 20

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

$\square$

Reference

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