Q1 $E(\text{Score}) = 4 \cdot P(\text{score}=4) + 3 \cdot P(\text{score}=3) + 2 \cdot P(\text{Score}=2) + 0.$

$4 \cdot P(\text{score}=4) = 4\int_0^{\frac{1}{\sqrt{3}}} f(x)\,dx = 4 \cdot \frac{1}{3} = \frac{4}{3}$

$3 \cdot P(\text{score}=3) = 3 \int_{\frac{1}{\sqrt{3}}}^{1} f(x)\,dx = 3 \cdot \frac{1}{6} = \frac{1}{2}.$

$2 \cdot P(\text{Score}=2) = 2 \cdot \int_1^{\sqrt{3}} f(x)\,dx = 2 \cdot \frac{1}{6} = \frac{1}{3}.$

therefore

Expected score of Single shot

$= \frac{4}{3} + \frac{1}{2} + \frac{1}{3} = \frac{8+3+2}{6} = \frac{13}{6}$

Q2 n observations $x_1, x_2, x_3 \ldots\ldots x_n$ independent

$P(X_1=x_1, X_2=x_2 \ldots X_n=x_n) = P(X=x_1)P(X=x_2) \cdots P(X=x_n)$

$= \theta e^{-\theta x_1}\, \theta e^{-\theta x_2} \cdots \theta e^{-\theta x_n}$

Want to Maximize this value. $\theta^n e^{-\theta(\sum_{i=1}^{n} x_i)}$

$$\left[\theta^n e^{-\theta(\sum_{i=0}^{n} x_i)}\right]' = n\theta^{n-1} \cancel{e^{-\theta(\sum_{i=1}^{n} x_i)}} + \theta^n \cancel{e^{-\theta(\sum_{i=1}^{n} x_i)}} \cdot \left(-\sum_{i=1}^{n} x_i\right) = 0$$

~~[illegible]~~ $n = \theta \sum_{i=1}^{n} x_i$ $\quad \theta = \frac{n}{\sum_{i=1}^{n} x_i}$

Q3 a) $X^TAX = [x_1 \cdots x_n] \begin{bmatrix} \sum a_{1i}x_i \\ \sum a_{2i}x_i \\ \vdots \\ \sum a_{ni}x_i \end{bmatrix}$

$= x_1 \sum_{i=1}^{n} a_{1i}x_i + x_2 \sum_{i=1}^{n} a_{2i}x_i \cdots + x_n \sum_{i=1}^{n} a_{ni}x_i$

$= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ji}x_i x_j$

b) Symmetric then $a_{ij} = a_{ji}$

plug in $x = e_1, e_2, e_3 \cdots$

$x_i = e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{bmatrix} \rightarrow i\text{th}$

$x_i^T A x_i = a_{ii} > 0$ according to the definition of positive matrix

Q4: a). A positive $X^TAX>0$

~~A+γI~~ $X^T(A+\gamma I)X = X^TAX + X^T\gamma I X = X^TAX + \gamma X^TX$

from difinition. $X^TAX>0$ and γX^TX also >0.

then $A+\gamma I$ also positive.

b). V eigenvector and λ the corresponding eigenvalue.

then $Av=\lambda v$.

$V^TAV = V^T\lambda V = \lambda V^TV > 0$. $\Leftarrow$ $V^TAV>0$

and $V^TV>0$ then $\lambda>0$

$\therefore$ all eigenvalue of A greater than zero

c). A positive $\Rightarrow$ $\forall x \in R^n$ $X^TAX>0$.

then Null space of A is ϕ

Because if $\exists X$ $AX=0$ then $X^TAX=0$ (X)

Null space $=0$ A has n independent column span R^n

thus invertible.

d). A Symmetric then A can be written as $U^T\Lambda U$.

U^T constructs by ~~eigenvalues~~ normalized eigenvectors and Λ is diagonal, eigenvalues.

assume $v_1, v_2, v_3, \dots v_n$ are n normalized eigenvector

$\lambda_1, \lambda_2, \lambda_3 \dots \lambda_n$ are n eigenvalues all of them are positive

Λ diagonal and all entries >0

then
$$\begin{vmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \lambda_3 & \\ 0 & & & \ddots \ \lambda_n \end{vmatrix} = \begin{vmatrix} \sqrt{\lambda_1} & & & 0 \\ & \sqrt{\lambda_2} & & \\ & & \sqrt{\lambda_3} & \\ 0 & & & \ddots \ \sqrt{\lambda_n} \end{vmatrix}^2 = (\sqrt{\Lambda})^2$$

therefore $A = U^T\sqrt{\Lambda}^T\sqrt{\Lambda}U$

$= (\sqrt{\Lambda}U)^T(\sqrt{\Lambda}U)$.

Q5.

a). $X, a \in \mathbb{R}^n$ get $\frac{\partial(X^Ta)}{\partial X}$

$f(X) = X^Ta = x_1a_1 + x_2a_2 + x_3a_3 + x_4a_4 \cdots\cdots$

$$\frac{\partial(X^Ta)}{\partial X} = \begin{vmatrix} \frac{\partial f(x)}{x_1} \\ \frac{\partial(f(x))}{x_2} \\ \vdots \\ \vdots \end{vmatrix}^T = a^T.$$

$$(a^T)^T = a$$

b). $A \in \mathbb{R}^{n \times n}$ $X \in \mathbb{R}^n$ $\frac{\partial(X^TAX)}{\partial X}$

$$AX = \begin{vmatrix} \sum A_{1i}X_i \\ \sum A_{2i}X_i \\ \sum A_{3i}X_i \\ \vdots \\ \vdots \end{vmatrix}$$

$f(X) = X^TAX = x_1\sum A_{1i}X_i + x_2\sum A_{2i}X_i \cdots\cdots + X_n\sum A_{ni}X_i$

$$\frac{\partial f(x)}{\partial x_k} = x_1A_{1k} + x_2A_{2k} + \cdots \cancel{x_k} \cdots + x_nA_{nk}$$

$$\cancel{\sum A_{ki}} \quad \sum_{i \neq k} A_{ki}X_i + 2A_{kk}X_k$$

$$= \sum_{i=1}^{n} A_{ki}X_i + \sum_{i=1}^{n} A_{ik}X_i$$

$$\frac{\partial f(x)}{\partial X} = \left| \left(\sum_{i=1}^{n} A_{1i}X_i + \sum_{i=1}^{n} A_{i1}X_i\right), \left(\sum_{i=1}^{n} A_{2i}X_i + \sum_{i=1}^{n} A_{i2}X_2\right) \cdots\cdots \right|$$

$$= X^TA^T + X^TA$$

$$(X^TA^T + X^TA)^T = AX + A^TX$$

c). $(XA)_{ij} = \sum_{k=1}^{n} X_{ik} A_{kj}$ Trace $(XA) = \sum_{i=1}^{n} (XA)_{ii}$

$$= \sum_{i=1}^{n} \sum_{k=1}^{n} X_{ik} A_{ki}$$

$$\frac{\partial (XA)}{\partial X_{ab}} = A_{ba} \quad \frac{\partial (XA)}{\partial X} = \begin{vmatrix} A_{11} & A_{21} & A_{31} & \cdots \\ A_{12} & & & \\ \vdots & & & \end{vmatrix} = A^T$$

d). want to proof

$$\sqrt{\sum X_i^2} \le \sum |X_i| \le \sqrt{n} \sqrt{\sum X_i^2}$$

$$\left(\sqrt{\sum X_i^2}\right)^2 \le \left(\sum |X_i|\right)^2$$

$$\downarrow$$

$$\sum X_i^2 \le \sum |X_i|^2 + \underbrace{2 \sum_{\substack{i=1 \\ j=1}}^{n} |X_i| \cdot |X_j|}_{\text{greater or equal to zero}} \quad i \ne j$$

therefore

$\|X\|_2 \le \|X\|_1$

want to show $\sum |X_i| \le \sqrt{n \sum |X_i|^2}$.

$$\left(\sum |X_i| \cdot 1\right)^2 \le \sum |X_i|^2 \sum^{n} (1^2)$$ Apply Cauchy inequality

$$\sum a_i^2 \sum b_i^2 \ge \left(\sum a_i b_i\right)^2$$

$$\sum |X_i|^2 \sum 1^2 \ge \left(\sum |X_i| \cdot 1\right)^2.$$ Done

e). Maximize $g(x) = X^T Z \;\big|\; \|Z\|_1 \le 1$

$$X^T Z = X_1 Z_1 + X_2 Z_2 + \cdots X_n Z_n \le \operatorname{Max}(|X_i|)(Z_1 + Z_2 + \cdots Z_n)$$

$$\le \operatorname{Max}(|X_i|) \cdot 1$$

Q6 a). $P(X|W_i) = N(\mu_i \sigma^2)(X)$. $\quad P(W_i|X) = \dfrac{P(X|W_i)P(W_i)}{P(X)}$

$$P(X) = P(X|W_1)P(W_1) + P(X|W_2)P(W_2)$$
$$= \tfrac{1}{2}N(\mu_1 \sigma^2)(X) + N(\mu_2, \sigma^2)(X)\tfrac{1}{2}$$

$$= \frac{\frac{1}{2} \cdot N(\mu_1, \sigma^2)(X)}{\frac{1}{2}\left(N(\mu_1 \sigma^2)(X) + N(\mu_2, \sigma^2)(X)\right)}$$

$$P(W_1|X) = \frac{\frac{1}{2}N(\mu_1 \sigma^2)(X)}{\frac{1}{2}[N(\mu_1 \sigma^2) + N(\mu_2, \sigma^2)](X)} \qquad = P(W_2|X) = \frac{\frac{1}{2}N(\mu_2 \sigma^2)(X)}{\frac{1}{2}[N(\mu_1 \sigma^2) + N(\mu_2 \sigma^2)](X)}$$

$\mu_1 \leq \mu_2$

$$N(\mu_1 \sigma^2)(X) = N(\mu_2 \sigma^2)(X)$$

$$\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(X-\mu_1)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}}$$

$$(x-\mu_1)^2 = (x-\mu_2)^2 \qquad x_0 = \frac{\mu_1+\mu_2}{2}$$

Boundary $\quad x_0 = \dfrac{\mu_1+\mu_2}{2}$. $\quad$ when $x > x_0$ $\quad \omega_2$

$\quad$ else $\quad \omega_1$

b) $P(\text{error}) = \underbrace{P(\text{misclassified as } w_2 | w_1) P(w_1)}_{A} + \underbrace{P(\text{misclassified as } w_1 | w_2) P(w_2)}_{B}$

$x_c = \frac{u_1+u_2}{2}$.

$$A: \frac{1}{2} \times \int_{x_c}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-u_1)^2}{2\sigma^2}} dx$$

$$B: \frac{1}{2} \cdot \int_{-\infty}^{x_c} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-u_2)^2}{2\sigma^2}} dx$$

$z_1 = \left(\frac{x-u_1}{\sigma}\right)$ replace x in A as z_1

$$A \Rightarrow \frac{1}{2} \int_{\frac{x_c-u_1}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$z_2 = \left(\frac{x-u_2}{\sigma}\right)$ replace x in B as z_2

$$B \Rightarrow \frac{1}{2} \int_{-\infty}^{\frac{x-u_2}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$A + B = P(\text{error})$ $\quad \cancel{A = \frac{1}{2} \int \frac{(x_c-u_1)}{\sigma}}$

$$B = \frac{1}{2} \int_{-\frac{(x-u_2)}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$ since $e^{-z_2^2/2}$ is even function

$= A$

therefore

$$P(\text{error}) = \int_{\frac{u_2-u_1}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

Q7. a). $f(w) = \frac{1}{2}\|Xw - y\|_2^2 + \frac{\lambda}{2}\|w\|_2^2$

$= \frac{1}{2}(Xw-y)^T(Xw-y) + \frac{\lambda}{2}w^Tw$

$\frac{\partial f(w)}{w}$ $(Xw-y) = k_{(w)}$ $\frac{\partial k^Tk}{\partial k}\frac{\partial k}{\partial w} = 2k^T \cdot X$

therefore $\frac{\partial f(w)}{w} = \frac{1}{2}\cdot 2(Xw-y)^T\cdot X + \frac{\lambda}{2}\cdot 2w^T$

$= (Xw-y)^T X + \lambda w^T$

b). want derivative $= 0 = \lambda w^T + (Xw-y)^T X$

$= \lambda w^T + w^T X^T X - y^T X$

$y^T X = w^T(\lambda I + X^T X)$.

$X^T y = (\lambda I + X^T X)^T w$.

$X^T y = (\lambda I + X^T X)\cdot w$.

since $\lambda > 0$ therefore $(\lambda I + X^T X) \neq 0$ ~~symmetric~~ invertable

$\Rightarrow (\lambda I + X^T X)^{-1} X^T y = w$.

c).

the first term

$\frac{1}{2}\|Xw-y\|_2^2 = \frac{1}{2}(Xw-y)^T(Xw-y)$. $w = w_n + X^T a$.

$Xw = Xw_n + XX^T a$ since $w_n \in Null(X)$

$Xw_n = 0$

$Xw = XX^T a$.

therefore $\frac{1}{2}\|Xw-y\|_2^2 = \frac{1}{2}(XX^T a - y)^T(XX^T a - y)$

only depends on a.

d). $\frac{\lambda}{2}w^Tw$ $w = w_n + X^T a$

$\frac{\lambda}{2}(w_n^T + a^T X)(w_n + X^T a)$

$= \frac{\lambda}{2}(w_n^T w_n + \underline{w_n^T X^T a} + \underline{a^T X w_n} + a^T X X^T a)$.

$w_n^T X^T a = w_n \cdot X^T a$

$a^T X w_n = (X^T a)^T \cdot w_n$

$= X^T a \cdot w_n = 0$

therefore $\frac{\lambda}{2}\|W\|_2^2 = \frac{\lambda}{2}[\underbrace{W_n^TW_n}_{\geq 0} + \underbrace{(X^Ta)^T(X^Ta)}_{\geq 0}]$

if a fixed $\frac{\lambda}{2}\|W\|_2^2 \geq \frac{\lambda}{2}(X^Ta)^T(X^Ta)$.

equal when $W_n = 0$

e7. minimize $\frac{1}{2}\|XX^Ta - y\|_2^2 + \frac{\lambda}{2}\|X^Ta\|_2^2$ ⇒ depend only on a

since from c and d. We find that

the first term of (1) not depend on W_n and the second therm of (1) is minimized (over W_n) when $W_n = 0$.

f7. then take derivative of $\frac{1}{2}\|XX^Ta - y\|_2^2 + \frac{\lambda}{2}\|X^Ta\|_2^2$. (over a)

$$\frac{\partial}{\partial a}\left[\frac{1}{2}(a^TXX^T - y^T)(XX^Ta - y) + \frac{\lambda}{2}(a^TX)(X^Ta)\right]$$

$$= \frac{\partial}{\partial a}\left[\frac{1}{2}(a^TXX^TXX^Ta - a^TXX^Ty - y^TXX^Ta + y^Ty) + \frac{\lambda}{2}a^TXX^Ta\right]$$

$$= \frac{1}{2}(a^TXX^TXX^T + a^TXX^TXX^T - y^TX^TX - y^TXX^T) + \frac{\lambda}{2\cdot}a^TXX^T]$$

$XX^T = A$ Asymmetric

then $a^TA^2 - y^TA + \lambda a^TA = 0$

$a^T(A^2 + \lambda A) = y^TA$ A invertiable

$(A^TA + \lambda A)a = Ay$

$A^{-1}(AA^T - A\lambda)a = A^{-1}Ay$

⇒ $(A^T + \lambda I)a = y$ $A^T = XX^T$ semipositive.

By Q4. part a. $v^TXX^Tv = \|X^Tv\|_2^2 \geq 0$.

$(A^T + \lambda I)$ $\lambda > 0$ invertiable

since. $A^T = XX^T$ semipositive therefore $a = (A^T + \lambda I)^{-1}y$

$= (XX^T + \lambda I)^{-1}y$

#7

q7. find $a_* : (XX^T + \lambda I)^{-1} y$.

$X = n \times d$.

get XX^T takes $n \times n \times d$.

inverse takes n^3.

$y : n \times 1$

then takes n^2

$O(n^3 + n^2 + n^2 d)$.

$= O(n^3 + n^2 d)$ if n.d huge.

find w_* $(\lambda I - X^T X)^{-1} X^T y$

X^T $d \times n$

$X^T X$ takes $d^2 n$

inverse takes d^3.

$X^T y$ takes $d \times n$

multiply $\underset{d \times d}{(\lambda I - X^T X)^{-1}}$ and $\underset{d \times 1}{X^T y}$

takes d^2

therefore $O(d^2 + nd + d^3 + d^2 n)$ if d, n large

$= O(d^3 + d^2 n)$

Compare

$a_* : O(n^3 + n^2 d)$ and $w_* : O(d^3 + d^2 n)$

if $n \gg d$ should choose w_*

if $n \ll d$ a_*

Q8 a7. W^*

$$\sum_{0}^{n} (x_i^T W - y_i^T)(W^T x_i - y_i) + \lambda \underbrace{\sum_{j=1}^{k}\sum_{i=1}^{d} W_{ij}^2}_{= tr(W^T W)}$$

$$= \sum_{0}^{n} \left[x_i^T W W^T x_i + x_i^T W y_i - y_i^T W^T x_i + y_i^T y_i\right] + tr(W^T W)$$

take derivative (over w). $\Rightarrow$

$$= \left\{ \sum_{i=\phi}^{n} \left[\cancel{2W^T x_i x_i^T} + 2W^T x_i x_i^T - 2 y_i x_i^T \right] + 2\lambda W^T \right\}^T$$

$$= \left\{ 2W^T \sum_{i=1}^{n} x_i x_i^T - 2 \sum y_i x_i^T + 2\lambda W^T \right\}^T$$

$$\sum \underset{d\times 1}{x_i}\ \underset{1\times d}{x_i^T} = \underset{d\times d}{\sum x_{i1} x_{i1} + \sum x_{i1} x_{i2}} \quad \sum x_{i1} x_{in}$$

$$\sum x_{i2} x_{i1} \cdots$$

$$= X X^T \quad X = [x_1, x_2, x_3, \cdots] \quad d\times n$$

$$\sum Y_i x_i^T = Y X \quad Y = [Y_1\ Y_2\ Y_3 \cdots Y_n] \quad k\times n$$

therefore gradient $= \left\{ 2W^T X X^T - 2 Y X^T + 2\lambda W^T \right\}$

want gradient $= 0$ then $2W^T X X^T + 2\lambda W^T = 2 Y X^T$

$$W^T (2XX^T + 2\lambda I) = 2 Y X^T$$

if $\lambda > 0$ $(2XX^T + 2\lambda I)$ positive

thus invertiable.

$$(2XX^T + 2\lambda I)\ W = 2XY^T$$

$$W = (2XX^T + 2\lambda I)^{-1} \cdot 2XY^T$$

$$= (XX^T + \lambda I)^{-1} \cdot XY^T$$

b) > python code
c)

d) to compute $\underset{d\times d}{(XX^T + \lambda I)^{-1}} \cdot \underset{dk}{XY^T}$ $\to n\times k$

$X: d\times n$

XX^T takes $d^2 n$

inverse takes d^3.

XY^T takes dnk

~~therefore $O(d^2n + d^3 + dnk)$~~

and $(XX^T + \lambda I)^{-1} \cdot XY^T$

takes d^2k

therefore $O(d^2k + dnk + d^3 + d^2n)$

e) $W = [w_1, w_2, \cdots w_k]$. $X: d\times n$

original classifier $\sum_{i=0}^{n} \|W^T x_i - y_i\|_2^2 + \lambda \|W\|_F^2$

$$W^T x_i = \begin{vmatrix} w_1 \cdot x_i \\ w_2 \cdot x_i \\ \vdots \\ w_k \cdot x_i \end{vmatrix} \qquad \|W\|_F^2 = \|w_1\|_2^2 + \|w_2\|_2^2 + \|w_3\|^2 + \cdots \|w_k\|^2$$

$$X = [x_1, x_2, \cdots x_n]$$

so we can choose w_i orthogonal to all x_i

then $X^T w_i = 0$.

and we can take w apart $W_\perp$ and $W_\parallel$

$W_\perp$ is consisted by w_i $X^T w_i = 0$

$W_\parallel$ is consisted by $w_i \in$ column space of X.

then $W = W_\perp + XA$ $\qquad A: n\times k$ matrix

We then plug this w in the equation.

$W = W_\perp + XA$

the first term

$W_\perp^T x_i = 0$

$$\sum_{i=0}^{n} \|(W_\perp^T + A^T X^T) x_i - y_i\|_2^2 = \sum_{i=0}^{n} \| A^T X^T x_i - y_i \|_2^2$$

$$= \sum_{i=1}^{n} (x_i^T X A A^T X^T x_i - 2 x_i^T X A y_i + y_i^T y_i).$$

$$= \mathrm{trace}(X^T X A A^T X^T X) - 2\mathrm{trace}[X^T X A Y] + \underline{\mathrm{trace}(Y^T Y)}$$

not about A

tail: $\|W\|_F^2 = \|W_\perp + XA\|_F^2 = \mathrm{trace}(W_\perp + XA)^T (W_\perp + XA)$

$= \mathrm{trace}(A^T X^T X A)$

$(W_\perp^T + A^T X^T)(W_\perp + XA) = \cancel{W_\perp^T W_\perp} + A^T X^T W_\perp$ 0

$+ \cancel{W_\perp^T W_\perp} + A^T X^T X A$

the first term does not depend on $W_\perp$

and for the second term: when $W_\perp = 0$ has minimum value

We take derivative

$$\frac{\partial}{\partial A} \left(\mathrm{trace}(X^T X A A^T X^T X) - 2\mathrm{trace}[X^T X A Y] + \lambda \mathrm{trace}(A^T X^T X A) \right) = 0$$

then $\lambda \cdot 2A^T \underbrace{X^T X} + 2A^T X^T X \underbrace{X^T X} = 2Y \underbrace{X^T X}$ Assume $X^T X$ invertiable

then $\lambda \cdot 2A^T + 2A^T X^T X = 2Y$.

$A^T(\lambda I + X^T X) = Y$ Since $(\lambda I + X^T X)$ invertiable

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

complexity: $X^T X$ takes $n^2 d$ inverse n^3 times Y^T $n^2 k$.

$O(n^2 d + n^3 + n^2 k)$. if $n \ll d$ compute A instead of W.

~/Desktop/cs189/hw1_code/hw1.py UNREGISTERED

hw1.py

```python
from mnist import MNIST
import sklearn.metrics as metrics
import numpy as np
from numpy.linalg import inv

NUM_CLASSES = 10

def load_dataset():
    mndata = MNIST('./data/')
    X_train, labels_train = map(np.array, mndata.load_training())
    X_test, labels_test = map(np.array, mndata.load_testing())
    X_train = X_train/255.0
    X_test = X_test/255.0
    return (X_train, labels_train), (X_test, labels_test)


def train(X_train, y_train, reg=0):
    ''' Build a model from X_train -> y_train '''
    #here involve a hyper parameter
    inverse = inv(np.dot(np.matrix.transpose(X_train), X_train) + 0.5*np.identity(784))
    return np.dot(inverse, np.dot(np.matrix.transpose(X_train), one_hot(y_train)))
def one_hot(labels_train):
    '''Convert categorical labels 0,1,2,....9 to standard basis vectors in R^{10} '''
    return np.array([[1 if i == labels_train[k] else 0 for i in range(10)] for k in range(len(labels_train))])

def predict(model, X):
    ''' From model and data points, output prediction vectors '''
    result = np.dot(np.matrix.transpose(model), np.matrix.transpose(X)) #get a vector
    return [np.argmax(i) for i in np.matrix.transpose(result)]#single array with dim = 1*60000

if __name__ == "__main__":
    (X_train, labels_train), (X_test, labels_test) = load_dataset()
    model = train(X_train, labels_train)
    y_train = one_hot(labels_train)
    y_test = one_hot(labels_test)

    pred_labels_train = predict(model, X_train)
    pred_labels_test = predict(model, X_test)


    print("Train accuracy: {0}".format(metrics.accuracy_score(labels_train, pred_labels_train)))
    print("Test accuracy: {0}".format(metrics.accuracy_score(labels_test, pred_labels_test)))
```

Line 25, Column 1 Spaces: 4 Python