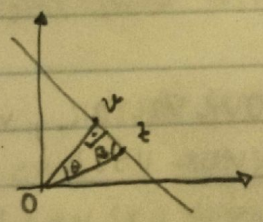
HOMEWORK 1 - ML FOR TEXT MINING (11641)
LEONARDO NEVES
ANDREW ID: LNEVES

1) A) WE KNOW THE NORMAL VECTOR FOR w IS $\frac{w}{\|w\|}$. ALSO, THE DISTANCE FROM THE PLANE TO THE ORIGIN HAVING A $90°$ ANGLE WITH H IS $d = (x-(0,0))\frac{w}{\|w\|}$. SINCE WE HAVE $w^T x = b$, WE CAN REWRITE $d = \frac{|b|}{\|w\|}$ AS THE DISTANCE FROM THE ORIGIN TO THE POINT v IN THE PLANE. LET'S NOW CONSIDER THE RIGHT TRIANGLE Δovz. THE ANGLE $o\hat{v}z = 90°$, MAKING $\theta + \beta = 90°$. ALSO, WE CAN WRITE, USING THE PYTHAGOREAN THEOREM, $\overline{ov}^2 + \overline{vz}^2 = \overline{oz}^2$, MAKING THE SEGMENT $\overline{ov}$ TO BE SMALLER THAN $\overline{oz}$ FOR ANY $z \neq v$. SINCE $\overline{ov} = \frac{|b|}{\|w\|}$, WE HAVE SHOWN THAT $\frac{|b|}{\|w\|}$ IS THE SHORTEST DISTANCE FROM O TO THE PLANE.



B) $h = f(x) = w^T x - b = 0$

WE SELECT THE PERPENDICULAR LINE FROM h TO x. WE KNOW A PERPENDICULAR LINE WILL HAVE THE FORM $w_2 x_1 - w_1 x_2 + c = 0$. IF THE POINT x IS DEFINED AS $x = (x_1', x_2')$, WE HAVE $c = -(w_2 x_1' - w_1 x_2')$. SO, THE EQUATION FOR THE LINE PERPENDICULAR TO h AND THAT PASSES THROUGH x IS

$$w_2 x_1 - w_1 x_2 - (w_2 x_1' - w_1 x_2') = 0$$

WE HAVE

$$w_1 x_1 + w_2 x_2 = b$$
$$w_2(x_1 - x_1') - w_1(-x_2 + x_2') = 0$$

SOLVING THIS SYSTEM WE FIND THE INTERSECTION POINT $z = (x_1, x_2)$

$$x_1 = (w_2^2 \cdot x_1' - w_1 w_2 x_2' + w_1 b) / (w_1^2 + w_2^2)$$
$$x_2 = (w_1^2 \cdot x_2' - w_1 w_2 x_1' + w_2 b) / (w_1^2 + w_2^2)$$

Scanned by CamScanner

WE KNOW THAT THE DISTANCE OF THE POINTS x AND z IS

$$d = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2}$$

BY SUBSTITUTING THE VALUES FOR x_1 AND x_2 IN d WE GET

$$d = \sqrt{\frac{w_2^2 x_1' - w_1 w_2 x_2' + w_1 b - x_1'(w_1^2 + w_2^2)}{w_1^2 + w_2^2} + \frac{w_1^2 x_2' - w_1 w_2 x_1' + w_2 b - x_2'(w_1^2 + w_2^2)}{w_1^2 + w_2^2}}$$

BY SIMPLIFYING (ALGEBRA OMITTED) WE GET

$$d = \sqrt{\frac{(w_1 x_1' + w_2 x_2' - b)^2}{w_1^2 + w_2^2}} = \frac{w_1 x_1' + w_2 x_2' - b}{\sqrt{w_1^2 + w_2^2}} = \frac{w^T x - b}{\|w\|}$$

SINCE WE WANT THE DIRECTION OF OUR VECTOR, WE INTRODUCE THE VARIABLE y, THE UNITARY VECTOR USED AS THE INDICATOR OF WHICH SIDE x IS LOCATED. THUS:

$$d = \frac{y(w^T x - b)}{\|w\|}$$

2) A)

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \quad p(\lambda) = \det \begin{bmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{bmatrix} = (3-\lambda)^2 - 1 = (\lambda - 2)(\lambda - 4)$$

$\lambda_1 = 2 \quad \lambda_2 = 4$

EIGENVALUE $\lambda_1 = 2$

$$\begin{bmatrix} 3-2 & 1 \\ 1 & 3-2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{FOR } \lambda_1 = 2,\ v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\begin{matrix} v_1 + v_2 = 0 \\ v_1 + v_2 = 0 \end{matrix} \Rightarrow v_1 = -v_2$$

Scanned by CamScanner

$\lambda_2 = 4$

$$\begin{bmatrix} 3-4 & 1 \\ 1 & 3-4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

FOR $\lambda_2 = 4$, $v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$\begin{matrix} -v_1 + v_2 = 0 \\ v_1 - v_2 = 0 \end{matrix} \Rightarrow v_1 = v_2$$

B) WE KNOW THAT, SINCE λ IS AN EIGENVALUE OF A,

$$Av = \lambda v \quad \text{for } v \neq 0$$

MULTIPLYING BOTH SIDES BY A:

$$A^2 v = A\lambda v = \lambda(Av) = \lambda(\lambda v) = \lambda^2 v$$

WE CAN SEE THAT, FOR A^2, THE EIGENVALUES ARE $\lambda_1^2, \ldots, \lambda_m^2$.

IF WE REPEAT THIS PROCESS m TIMES

$$A^m v = A^{m-1} \lambda v = \lambda^m v$$

BY INDUCTION, WE CAN PROVE THE EIGENVALUES OF A^{m+1} ARE λ^{m+1} USING THE SAME IDEA.

$$A^m . A . v = \lambda^m . (A . v)$$
$$A^{m+1} v = \lambda^m \lambda v = \lambda^{m+1} v$$

SINCE v DIDN'T CHANGE, WE HAVE ALSO SHOWN THAT EACH EIGENVECTOR OF A IS ALSO AN EIGENVECTOR OF A^m.

3) A) USING THE BERNOULLI TRIAL FORMULA

$$L(p) = \binom{m}{p} p^k (1-p)^{m-k}$$

Scanned by CamScanner

BY COMPUTING THE LOG WE GET

$$\log L(p) = \log\binom{n}{k} + k\log p + (n-k)\log(1-p)$$

BY TAKING THE DERIVATIVE AND MAKING IT TO ZERO

$$\frac{\partial \log L(p)}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} = 0$$

$$\frac{k}{p} = \frac{n-k}{1-p} \qquad k - \cancel{kp} = np - \cancel{kp}$$

$$\hat{p} = \frac{k}{n}$$

B) USING THE MULTINOMIAL DISTRIBUTION FORMULA

$$L(p) = \binom{n}{n_1 \dots n_m} \prod_{j=1}^{m} p_j^{n_j}$$

We apply the log:

$$\log L(p) = \log\binom{n}{n_1 \dots n_m} + \sum_{j=1}^{m} n_j \log p_j$$

TO SOLVE FOR P, WE NEED TO MAXIMIZE THE FUNCTION WITH THE CONSTRAINT $\sum_{j=1}^{m} p_j = 1$. APPLYING LAGRANGE MULTIPLIERS

$$\log \hat{L}(p) = \log\binom{n}{n_1 \dots n_m} + \sum_{j=1}^{m} n_j \log p_j + \lambda\left(1 - \sum_{j}^{m} p_j\right)$$

$$\frac{\partial \log \hat{L}(p)}{\partial p} = \frac{n_j}{p_j} - \lambda = 0, \quad \lambda = \frac{n_j}{p_j}$$

$$\frac{\partial \log \hat{L}(p)}{\partial \lambda} = 1 - \sum_{j}^{m} p_j = 0, \quad \sum_{j}^{m} p_j = 1$$

Scanned by CamScanner

$$\lambda p_j = m_j \xrightarrow{\text{SUM ON BOTH SIDES}} \lambda \sum_j^m p_j = \sum_j^m m_j \rightarrow \lambda = m$$

So, $\hat{p}_j = \frac{m_j}{m}$, $\forall j = 1, 2, \dots, m$

4) A) $u = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1}$

$$\frac{du}{dx} = -(1+e^{-x})^{-2} \frac{d}{dx}(1+e^{-x})$$

$$= -\left(\frac{1}{1+e^{-x}}\right)^2 \cdot \left(\emptyset + e^{-x} \frac{d}{dx}(-x)\right) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1-1+e^{-x}}{(1+e^{-x})^2} = \frac{(1+e^{-x})}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2}$$

$$= \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}}\left(1 - \frac{1}{(1+e^{-x})}\right) = u(1-u)$$

B) WE CHANGE x TO HAVE $x_0 = 1$ TO MAKE THE RESULTS MORE READABLE. $z = w_0 x_0 + w_1 x_1 + \dots + w_m x_m$

$$\frac{\partial L}{\partial w_i} = \frac{\partial}{\partial w_i} y \ln\left(\frac{1}{1+e^{-z}}\right) + (1-y) \ln\left(1 - \frac{1}{1+e^{-z}}\right)$$

$$= \frac{\partial}{\partial w_i} y \ln(u) \text{ (I)} + \frac{\partial}{\partial w_i} (1-y) \ln(1-u) \text{ (II)}$$

(I) $y \frac{\partial \ln u}{\partial w_i} \overset{\text{CHAIN RULE}}{=} y \frac{\partial \ln u}{\partial u} \frac{\partial}{\partial w_i} \frac{1}{1+e^{-z}} = y \frac{1}{u} \cdot \frac{\partial}{\partial w_i} (1+e^{-z})^{-1}$

$$\overset{\text{CHAIN RULE}}{=} y \cdot \frac{1}{u} \cdot \left(-(1+e^{-z})^{-2}\right) \cdot \frac{\partial}{\partial w_i}(1+e^{-z})$$

$$= y \cdot \frac{1}{u} \cdot \left(-\frac{1}{(1+e^{-z})^2} \cdot x_i (\emptyset - e^{-z})\right) \text{ (III)}$$

$$\boxed{\frac{1}{u} = 1+e^{-z}} \Rightarrow y(1+e^{-z}) \frac{1}{(1+e^{-z})^2} \cdot x_i \cdot e^{-z} = y x_i \left(\frac{e^{-z}}{1+e^{-z}}\right)$$

Scanned by CamScanner

II $(1-y)\frac{\partial}{\partial w_i}(\ln(1-u))$

WE CALL $\boxed{k = 1-u}$

$$= (1-y)\frac{\partial \ln k}{\partial k}\frac{\partial}{\partial w_i}\left(1-\frac{1}{1+e^{-z}}\right)$$

$$= (1-y)\frac{1}{k}\cdot\cancel{\frac{\partial(1)}{\partial w_i}}^{0} - \frac{\partial}{\partial w_i}\frac{1}{1+e^{-z}}$$ FROM III

$$= (1-y)\left(\frac{1}{k}\right)\left(\frac{x_i e^{-z}}{(1+e^{-z})^2}\right) \qquad \boxed{\frac{1}{k} = \frac{1+e^{-z}}{e^{-z}}}$$

$$= (1-y)\left(\frac{1+e^{-z}}{e^{-z}}\right)\left(\frac{x_i e^{-z}}{(1+e^{-z})^2}\right) = (1-y)\left(\frac{x_i}{1+e^{-z}}\right)$$

$$\frac{\partial L}{\partial w_i} = \frac{x_i}{1+e^{-z}}\left(y e^{-z} + (1-y)\right)$$

c) $$\frac{\partial L}{\partial w_i} = x_i y\left(\frac{e^{-z}}{1+e^{-z}}\right) + x_i(1-y)\left(\frac{1}{1+e^{-z}}\right)$$

$$\frac{\partial L}{\partial w_i w_j} = \frac{\partial x_i}{\partial w_i w_j} y\left(\frac{e^{-z}}{1+e^{-z}}\right)^{\mathrm{I}} + \frac{\partial}{\partial w_i w_j} x_i(1-y)\left(\frac{1}{1+e^{-z}}\right)^{\mathrm{II}}$$

I USING THE QUOTIENT RULE

$$= \frac{x_i y \frac{\partial}{\partial w_i w_j}(e^{-z})(1+e^{-z}) - \frac{\partial}{\partial w_i w_j}(1+e^{-z})(e^{-z})}{(1+e^{-z})^2}$$

Scanned by CamScanner

$$= x_i y \left(\frac{-x_j e^{-z}(1+e^{-z}) - (-x_j e^{-z}) e^{-z}}{(1+e^{-z})^2} \right)$$

$$= x_i x_j y e^{-z} \left(\frac{-(1+e^{-z}) + e^{-z}}{(1+e^{-z})^2} \right) = -x_i x_j y \cdot \left(\frac{e^{-z}}{(1+e^{-z})^2} \right)$$

(II) $$\frac{\partial x_i (1-y)}{\partial w_i w_j} \left(\frac{1}{1+e^{-z}} \right) = x_i (1-y) \frac{\partial}{\partial w_i w_j} (1+e^{-z})^{-1}$$

CH.RULE

$$= x_i (1-y) \left(-\frac{1}{(1+e^{-z})^2} \right) \left(-x_j e^{-z} \right) = x_i x_j (1-y) \frac{e^{-z}}{(1+e^{-z})^2}$$

So, $$\frac{\partial^2 L}{\partial w_i w_j} = x_i x_j \frac{e^{-z}}{(1+e^{-z})^2} (1-2y)$$

D) THE LOG-LIKELIHOOD FORMULA IS

$$\log L = -\sum_i^N \log(1 + e^{-y_i w^T x_i})$$

The first derivative is

$$\frac{\partial \log L}{\partial w} = -\sum_i \left(\frac{1}{1+e^{-z}} - y_i \right) x_i$$

The Hessian is going to be, using $\mu = \frac{1}{1+e^{-z}}$

$$H = \frac{\partial^2 \log L}{\partial w} = -\sum_i \mu(1-\mu) x_i \cdot x_i^T = -X^T S X, \text{ where}$$

$S = \text{diag}(\mu(1-\mu))$. As we know $0 < y_m < 1$, we see that, for an arbitrary vector v, $v^T H v < 0$. From this we prove that the Hessian is negative semi-definite and, thus, the function is concave.

Scanned by CamScanner