ECE 146 Discussion Set 1
Introduction to Machine Learning
Instructor: Lara Dolecek TA: Zehui Chen

Alex chen1046@ucla.edu.

1. Linear Algebra Review

(a) What are the four properties of a inner product i.e. $< x, y >$?

x^Ty $\quad \|x\|_2^2 = 0 \iff x = 0$

1. Positivity $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0 \iff x = 0$ $\quad x^Tx = \|x\|_2^2 \geq 0$
2. Symmetry $\langle x, y\rangle = \langle y, x\rangle$ $\quad x^Ty = y^Tx$
3. Additivity $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$ $\quad (x+z)^Ty = x^Ty + z^Ty$
4. Homogenity $\langle rx, y\rangle = r\langle x, y\rangle$ $\quad (rx)^Ty = r \cdot x^Ty$

(b) What are the three properties of a norm $f(x)$?

1. Positivity $f(x) \geq 0$ $\quad f(x) = 0 \iff x = 0$
2. Homogenity $f(rx) = |r| f(x)$
3. Triangle Inequality $f(x + y) \leq f(x) + f(y)$

(c) What is the L2-norm? L1-norm? L∞-norm?

L2 norm: $\|x\|_2 = \sqrt{x^Tx} = \sqrt{\sum_i x_i^2}$ $\quad \|x\| = \|x\|_2$

L1 norm: $\|x\|_1 = \sum_i |x_i|$

L∞ norm: $\|x\|_\infty = \max_i |x_i|$

Normalization

(d) What does it mean to normalize a vector?

① Bounded norm $\quad x^{(i)} \leq 1$ $\quad$ E.g. GPA. GRE 0-4 0-340 ⇓ 0-4 normalized

$\|x^{(i)}\| \leq R$

$x^{(1)}, x^{(2)}, x^{(3)} \dots$

Find $\max_i \|x^{(i)}\|_2$

$x^{(i)} := x^{(i)} / \max_i \|x^{(i)}\|_2$

② Gaussian Normalization.

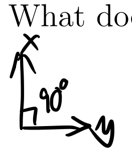
$x = [x_1 \ \dots \ x_n]^T$ $\quad \bar{x}$ $\quad$ std(x)

$\rightarrow N(0, 1)$ $\quad$ mean : 0 $\quad$ std : 1

$x_i := \frac{x_i - \bar{x}}{\text{std}(x)}$ $\leftarrow \alpha$, $\leftarrow \beta$

Batch normalization

(e) What does it mean for two vectors to be orthogonal?



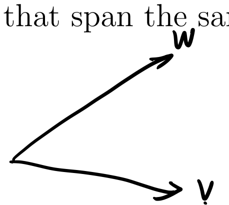
$x^Ty = 0$

orthonormal

$x^Ty = 0$

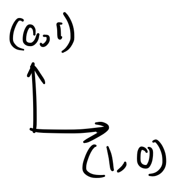
$\|x\| = \|y\| = 1$

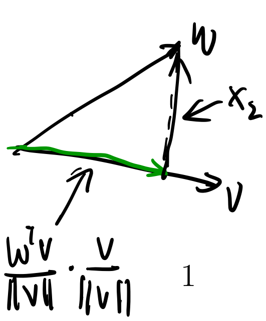
(f) Given two vectors **w** and **v** that are not orthogonal, find two orthogonal vectors that span the same space as **w** and **v**.



Gram-Schmidt Process. QR decomposition.

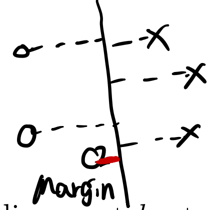
$x_1 = v / \|v\|$



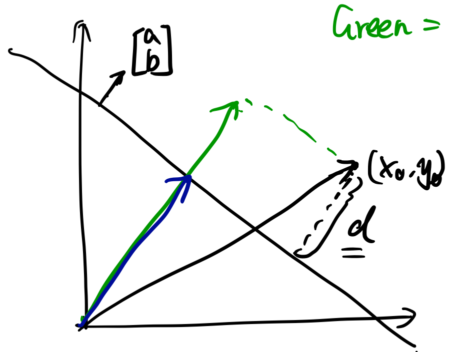


$x_2 = w - w^Tv \cdot \frac{v}{\|v\|^2}$

Perceptron Margin.



2. Given a line $ax + by + c = 0$ and a point (x_0, y_0), find the formula for the minimum distance between the point and the line.



$ax+by+c=0$

$[a\ b]\begin{bmatrix}x\\y\end{bmatrix} + c = 0$

Green $= \dfrac{[x_0\ y_0]\begin{bmatrix}a\\b\end{bmatrix}}{\sqrt{a^2+b^2}} \dfrac{\begin{bmatrix}a\\b\end{bmatrix}}{\sqrt{a^2+b^2}}$

Blue $= \begin{bmatrix}\alpha a\\ \alpha b\end{bmatrix} = -\begin{bmatrix}\frac{ac}{a^2+b^2}\\ \frac{bc}{a^2+b^2}\end{bmatrix}$

$\begin{bmatrix}\alpha a\\ \alpha b\end{bmatrix}^T \begin{bmatrix}x\\y\end{bmatrix} + c = 0$

$\Downarrow$

$\alpha = -\dfrac{c}{a^2+b^2}$

$d = \dfrac{\left| [x_0\ y_0]\begin{bmatrix}a\\b\end{bmatrix} + c \right|}{\sqrt{a^2+b^2}}$

$\Downarrow$

Generalize to $w^T x + b = 0$, x_0

$d = \dfrac{|x_0^T w + b|}{\|w\|_2}$

3. In class, you were taught to only consider the perceptron that goes through the origin. We will now show that this formulation is sufficient to encompass the case where the perceptron does not go through the origin.

   Consider the classification function of a perceptron classifier that does not go through the origin,

$$h(x) = w^T x + b$$

   where w and b are the hyper plane parameters.

   Now, consider the classification function of a perceptron classifier that does go through the origin

$$\tilde{h}(\tilde{x}) = \tilde{w}^T \tilde{x}.$$

   Find a way to formulate $\tilde{w}$ and $\tilde{x}$ in terms of x, b, w.

Explicit bias v.s. Implicit bias

Data: x — Data: $\tilde{x} = \begin{bmatrix}x\\1\end{bmatrix}$

Weight: w — Weight: $\tilde{w} = \begin{bmatrix}w\\b\end{bmatrix}$ $\quad \tilde{x}^T\tilde{w} = x^T w + b$

Bias: b

Perceptron in textbook — Perceptron in class

Support vector machine — Logistic Regression.

If $x^{(i)}$ is missclassified.
$w = w + y x^{(i)}$
$b = b + y$

If $x^{(i)}$ is missclassified
$\tilde{w} = \tilde{w} + y(\tilde{x}^{(i)})$

Label representation:

$y \in \{-1, 1\}$ $y \in \{0, 1\}$

Perceptron Logistic Regression.

Dataset. 2030.csv, 2031.csv, $y \in \{-1, 1\}$

$[0, 1] = [-1, 1]/2 + 0.5$

$[-1, 1] = 2 \times [0, 1] - 1$