① Recall convex hull of $x_1, \dots, x_n$ is $\{\alpha^T x \mid \alpha \in [0,1] \text{ and } \alpha^T \mathbb{1} = 1\}$

Lemma 1: Convex hulls are convex

pf: Let $\alpha^T x$, $\beta^T x \in$ convex hull.

$\hookrightarrow$ WTS: $t\alpha^T x + (1-t)\beta^T x \in$ convex hull $\forall\ 0 \le t \le 1$

$$= (t\alpha^T + (1-t)\beta^T)x$$

$$= (t\alpha + (1-t)\beta)^T x \quad \text{but } (t\alpha + (1-t)\beta)_n = t\alpha_n + (1-t)\beta_n \in [0,1]$$

$$(t\alpha + (1-t)\beta)^T \mathbb{1} = t\alpha^T \mathbb{1} + (1-t)\beta^T \mathbb{1} \ \checkmark$$

Lin. sep: $\exists\ \hat{w} \in \mathbb{R}^d$, $w_0 \in \mathbb{R}$ s.t.

$$\hat{w}^T x_i + w_0 < 0 \quad \forall\ i = 1, \dots, N$$

$$\hat{w}^T y_j + w_0 > 0 \quad \forall\ j = 1, \dots, M$$

Interesting Direction: suppose convex hulls of $x_1, \dots, x_N$ and $y_1, \dots, y_M$ do not intersect

WTS: $\exists\ \hat{w} \in \mathbb{R}^d$, $w_0 \in \mathbb{R}$ s.t. $\hat{w}^T x + w_0 < 0$, $\hat{w}^T y + w_0 > 0$ $\forall$ $x \in X$ = convex hull of x, $y \in Y$ = convex hull of y

Construction of $\hat{w}, w_0$:

Let $(x_0, y_0) = \underset{(x,y) \in X, Y}{\operatorname{argmin}} \|x - y\|$ $\Rightarrow$ Exists b/c X, Y are bounded + closed, so $X \times Y$ is compact. Since $\|x-y\|$ con't, it achieves a minimum.

Set $\hat{w} = \dfrac{y_0 - x_0}{\|y_0 - x_0\|}$ and $w_0 = -\left(\dfrac{y_0 + x_0}{2}\right)^T \hat{w}$

Since we only care about sign of $\hat{w}^T x + w_0$, we can instead use $\hat{w} = y_0 - x_0$ and $w_0 = -\dfrac{(y_0 + x_0)^T}{2} \hat{w}$

Let us now show the inequalities:

Let $x \in X$, and consider

$$w^T x + w_0 = (y_0 - x_0)^T x - \frac{(y_0 + x_0)^T (y_0 - x_0)}{2} \overset{\text{WTS}}{<} 0$$

$$\Rightarrow (y_0 - x_0)^T \left(x - \frac{y_0 + x_0}{2}\right) = \frac{1}{2}(y_0 - x_0)^T \left((x - y_0) + (x - y_0)\right)$$

Remains to show: $(y_0 - x_0)^T (x - y_0) < -(y_0 - x_0)^T (x - x_0)$

we will show $\to\ >0$

Showing $-(y_0-x_0)^T(x-x_0) > 0$

$\Rightarrow (y_0-x_0)^T(x-x_0) < 0$

Pf: Key step, consider: $\ell(t) = tx + (1-t)x_0$ for $0 \le t \le 1$

$(y_0-x_0)^T(\ell(t)-x_0)$ $\quad$ ↳ in X by convexity

observation: by def. $\|y_0 - x_0\| \le \|\ell(t) - y_0\|$ for all t

so $\|y_0-x_0\|^2 - \|\ell(t)-y_0\|^2 \le 0$

$(y_0-x_0)^T(y_0-x_0) - (\ell(t)-y_0)^T(\ell(t)-y_0) \le 0$

$(y_0-x_0)^T(y_0-x_0) - (tx + (1-t)x_0 - y_0)^T(tx + (1-t)x_0 - y_0) \le 0$

$(y_0-x_0)^T(y_0-x_0) - (tx - tx_0 + (x_0-y_0))^T(tx - tx + (x_0-y_0)) \le 0$

$-t^2(x-x_0)^T(x-x_0) - 2t(x-x_0)^T(x_0-y_0) \le 0$

when $t=0$ this is obviously true since $\ell(0) = x_0$ so,

$-t(x-x_0)^T(x-x_0) - 2(x-x_0)^T(x_0-y_0) < 0 \quad \forall\ 0 < t \le 1$

$\Rightarrow (x-x_0)^T(y_0-x_0) < \frac{1}{2}t(x-x_0)^T(x-x_0)$

Taking $t \to 0$ shows $(x-x_0)^T(y_0-x_0) < 0$

$\therefore$ If they are linearly seperable, their convex hulls do not intersect.

Other direction:

Consider x_0 in the convex hull of X, then

$\exists\ \alpha$ with $\sum_{i=1}^{n} a_n = 1$ and $\alpha_n \ge 0$ and $x_0 = \sum \alpha_n x_n$

$\exists\ \beta$ with $\sum \beta_n = 1$ and $\beta_n \ge 0$ and $x_0 = \sum \beta_n y_n$

where y_n represents points in hull of y.

Suppose then that they are linearly separable, meaning

$w^T x_0 + w_0 > 0$ and also $w^T x_0 + w_0 < 0$ for linear decision boundary defined by w and w_0. This is impossible which means that we see by contradiction that if the convex hulls intersect they are not lin sep.

Exercise 4.14:

Need to find $\underset{w}{\operatorname{argmax}} \log p(t_n | \phi(x_n))$

$$\Rightarrow p(C=0 | \phi(x_n))^{1-t_n} p(C=1 | \phi(x_n))^{t_n}$$

$$= \sigma(w^T \phi(x_n) + w_0)(1 - \sigma(w^T \phi(x_n) + w_0))$$

Since $\mathbb{P}(C=1 | \phi(x_n)) = 1 - \mathbb{P}(C=0 | \phi(x_n)) = 1 - \sigma(w^T \phi(x_n) + w_0)$

$$= 1 - \frac{1}{1 + e^{-w^T \phi(x_n) + w_0}}$$

$$= \frac{e^{-w^T \phi(x_n) - w_0}}{1 + e^{\underbrace{-w^T \phi(x_n) - w_0}_{\gamma}}} = \left(\frac{1}{1+e^{\gamma}}\right) \frac{e^{\gamma}}{1+e^{\gamma}}$$

$$\Rightarrow \underset{w}{\operatorname{argmax}} \log \prod_{n=1}^{N} \left(\frac{1}{1+e^{\gamma}}\right)^{t_n} \left(\frac{e^{\gamma}}{1+e^{\gamma}}\right)^{1-t_n}$$

$$\underset{w}{\operatorname{argmax}} \sum t_n \log\left(\frac{1}{1+e^{\gamma}}\right) + \sum (1-t_n) \log\left(\frac{e^{\gamma}}{1+e^{\gamma}}\right)$$

$$\underset{w}{\operatorname{argmax}} \sum -t_n \log(1+e^{\gamma}) - \gamma - \log(1+e^{\gamma})$$

$$\underset{w}{\operatorname{argmax}} \sum (t_n - 1)(w^T \phi(x_n) + w_0) - \log(1+e^{\gamma})$$

consider the 2 cases $t_n = 1$, $t_n = -1$:

if $t_n = 1 \Rightarrow \underset{w}{\operatorname{argmax}} \sum -\log(1+e^{\gamma}) = \log\left(\frac{1}{1+e^{\gamma}}\right)$

$\hookrightarrow$ this is maximised when $\|w\| \to \infty$

Similarly, if $t_n = -1 \Rightarrow \underset{w}{\operatorname{argmax}} -2(w^T \phi(x_n) + w_0) - \log(1+e^{\gamma})$

which is maximised when $\|w\| \to \infty$

---

github link: github.com/mateobianchi6/156