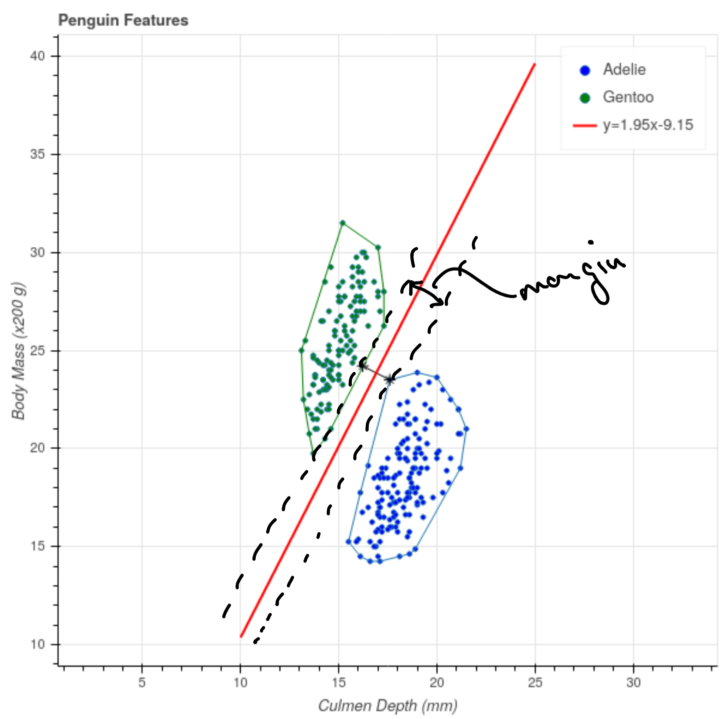
Optimal Margin and Closest Points

Convex Hulls and Margins

Our goal for this section is to reduce the optimal margin problem for $A^{\pm}$ to the problem of finding the closest points in the convex hulls $C(A^{\pm})$.



Solution

Optimal margin vs distance

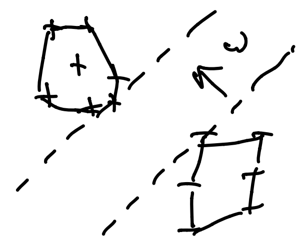
Proposition: The optimal margin between linearly separable sets $A^{\pm}$ is at most the distance between their convex hulls:

$$\underline{\tau(A^+, A^-)} \leq \min_{\substack{p \in C(A^+) \\ a \in C(A^-)}} \|p - q\|.$$

Proof:

- Choose a pair of functions $f^{\pm}(x) = w \cdot x - B^{\pm}$ so that $f^{\pm}(x) = 0$ are supporting hyperplanes for $A^{\pm}$ respectively.

- These are supporting hyperplanes for $C(A^{\pm})$ also.



- If p and q are points in $C(A^+)$ and $C(A^-)$ then $w \cdot p - B^+ \geq 0$ and $w \cdot q - B^- \leq 0$.

$p \in C(A^+)$, $f^+(p) = w \cdot p - B^+ \geq 0$

$q \in C(A^-)$ $f^-(q) = w \cdot q - B^- \leq 0.$

- So $\underline{w \cdot (p - q) \geq B^+ - B^- \geq} 0$.

$$\begin{array}{r} w \cdot p - B^+ \geq 0 \\ -w \cdot q + B^- \geq 0 \\ \hline w \cdot (p - q) \geq B^+ - B^- > 0 \end{array}$$

- Therefore

$$\|p - q\| \geq \boxed{\frac{B^+ - B^-}{\|w\|} = \tau_w(A^+, A^-)}$$

$\tau(A^+, A^-) = \max_{w} \tau_w(A^+, A^-)$

$w \cdot (p - q) = \|w\| \|p - q\| \cos\theta \quad |\cos\theta| < 1$

- This is true for all w.

More on optimal margin and distance

Two Corollaries:

1. $A^{\pm}$ are linearly separable if and only if the minimal distance between their convex hulls is greater than zero.

2. If we can find supporting hyperplanes $f^{\pm} = 0$ whose margin **equals** the minimal distance between the convex hulls, then those **must be at** the optimal margin.

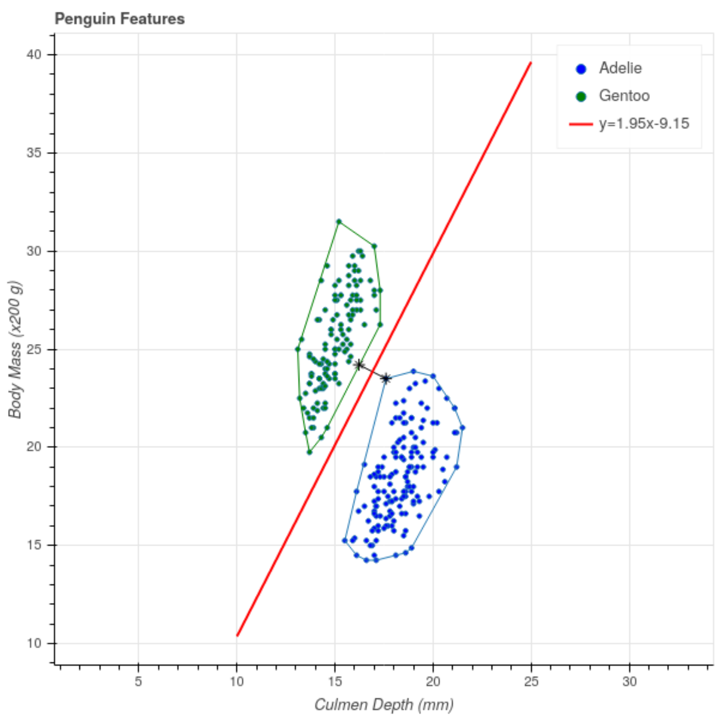
Convex Hulls and Margins

Proposition: Let A^+ and A^- be two linearly separable sets.

1. There are points $p^* \in C(A^+)$ and $q^* \in C(A^-)$ so that

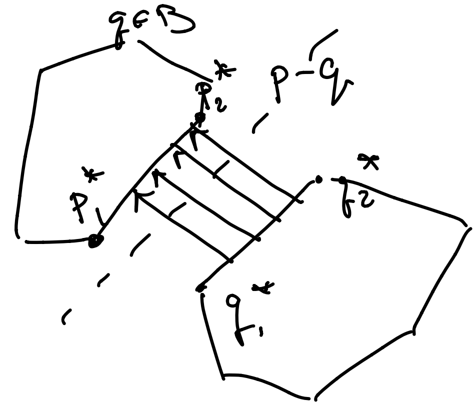
$$\|p^* - q^*\| = D = \min_{\substack{p \in C(A^+) \\ q \in C(A^-)}} \|p - q\|.$$

   Further, if (p_1^*, q_1^*) and (p_2^*, q_2^*) are two pairs of points with $D = \|p_1^* - q_1^*\| = \|p_2^* - q_2^*\|$ then $p_1^* - q_1^* = p_2^* - q_2^*$.



A: $[0 \le x < 1)$ B: $(2 < x \le 3]$

$$\min_{\substack{p \in A \\ q \in B}} (p - q) = 1$$



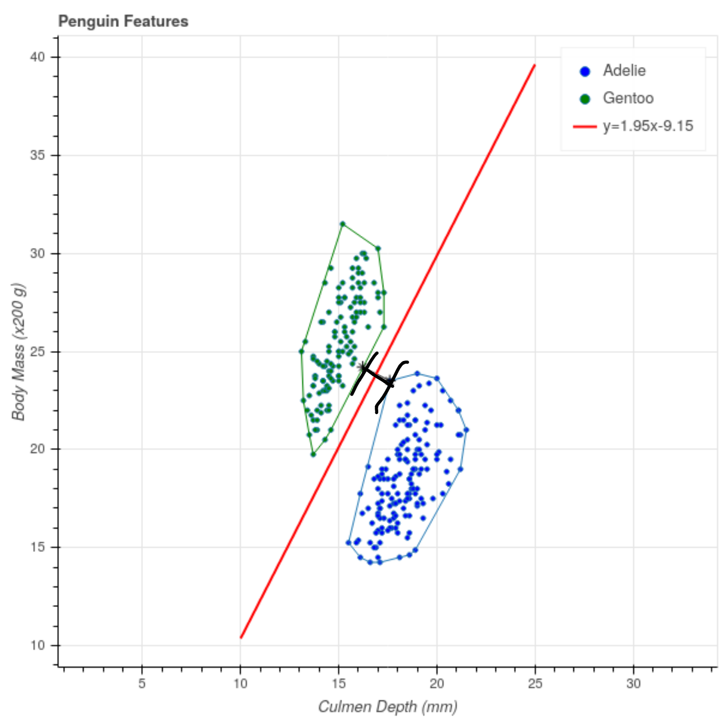
Convex Hulls and Margins

2. Let
   (a) $w = p^* - q^*$ ✓
   (b) $B^+ = w \cdot p^*$ ✓
   (c) $B^- = w \cdot q^*$. ✓
   (d) $f^{\pm}(x) = w \cdot x - B^{\pm}$

   Then the hyperplanes $f^{\pm}(x) = 0$ are supporting hyperplanes for $A^{\pm}$ respectively, and the associated margin

$$\tau_w(A^+, A^-) = \frac{B^+ - B^-}{\|w\|} = \|p^* - q^*\|$$

   is optimal.



Solution

Closest points

Proposition: Let

$$D = \min_{\substack{p \in C(A^+) \\ q \in C(A^-)}} \|p - q\|.$$

Then there exist points $p^* \in C(A^+)$ and $q^* \in C(A^-)$ so that $D = \|p^* - q^*\|$.

Suppose further that there are two pairs of points (p_1^*, q_1^*) and (p_2^*, q_2^*) in $C(A^+) \times C(A^-)$ with $D = \|p_1^* - q_1^*\| = \|p_2^* - q_2^*\|$. Then $p_1^* - q_1^* = p_2^* - q_2^*$.

Proof of the closest point proposition

The function $f(p,q) = \|p - q\|$ on $C(A^+) \times C(A^-)$ is a continuous function on a compact set, so it attains its minimum.

Suppose p_1^*, q_1^* and p_2^*, q_2^* have

$$D = \|p_1^* - q_1^*\| = \|p_2^* - q_2^*\| \Rightarrow \begin{array}{c} p_1^* - q_1^* \\ \shortparallel \\ p_2^* - q_2^* \end{array}$$

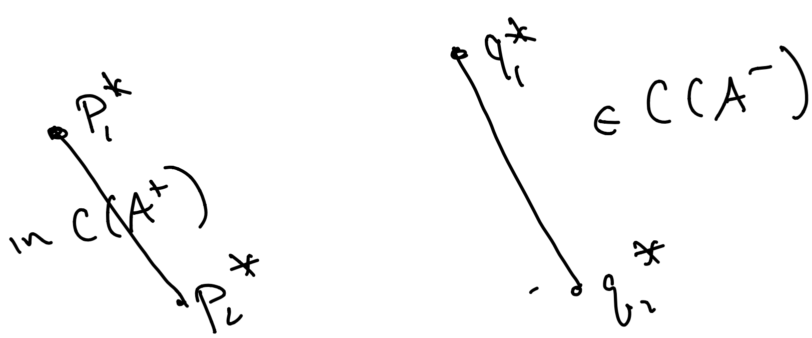
For $0 \le s \le 1$, let

$$p(s) = (1-s)p_1^* + sp_2^* \quad \in C(A^+)$$

and

$$q(s) = (1-s)q_1^* + sq_2^*. \quad \subseteq C(A^-)$$

These points are all in $C(A^+)$ and $C(A^-)$ respectively by convexity.

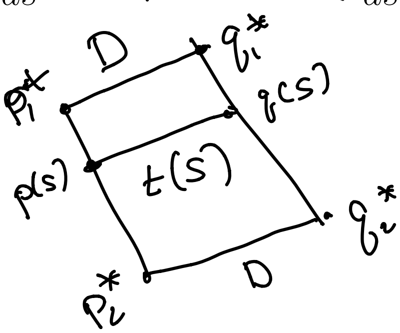


Let

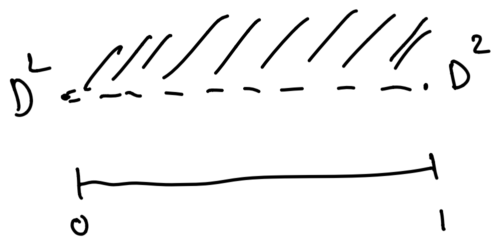
$$t(s) = \|p(s) - q(s)\|^2.$$

We must have $t(s) \geq D^2$ for all s. On the other hand

$$\frac{d}{ds}t(s) = 2(p(s)-q(s))\frac{d}{ds}(p(s)-q(s)) = 2(p(s)-q(s))\cdot((p_2^*-q_2^*)-(p_1^*-q_1^*))$$



$t(0) = t(1) = D^2$



Evaluate this at $s = 0$ and you get

$$\frac{d}{ds}t(s) = 2(p_1^* - q_1^*) \cdot ((p_2^* - q_2^*) - (p_1^* - q_1^*)) = 2(v_1 \cdot v_2 - \|v_1\|^2).$$

where $v_1 = p_1^* - q_1^*$ and $v_2 = p_2^* - q_2^*$.

Remember that $\|v_i\|^2 = D^2$ for $i = 1, 2$ and note that, as a result, $v_1 \cdot v_2 \le D^2$. Therefore

$$\frac{dt(s)}{ds}\Big|_{s=0} = 2(v_1 \cdot v_2 - \|v_1\|^2) \le 0$$

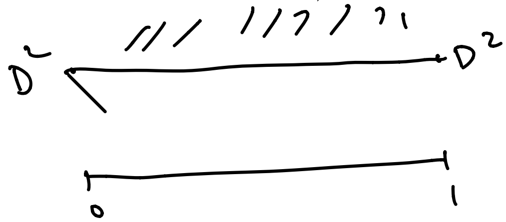
$\|$

D^2

$|v_1 \cdot v_2| = \|v_1\|\|v_2\| \cos\theta \le D^2$

$= D^2$ only if $v_1 = v_2$.

If the derivative of $t(s)$ were negative, $t(s)$ would be decreasing at $s = 0$ so there would be a point with $s > 0$ where $t(s) < D$. That can't happen, since D is the minimal distance, so $v_1 \cdot v_2 = D^2$ which means $p_1^* - q_1^* = p_2^* - q_2^*$.



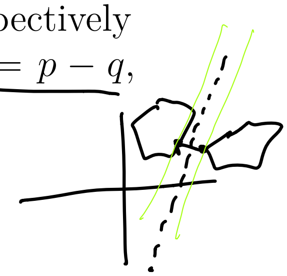
$t(s) < D^2$ can't happen.

$v_1 \cdot v_2 = D^2 \Rightarrow v_1 = v_2 = w$

$= p_1^* - q_1^* = p_2^* - q_2^*$

Closest points yield optimal margin

Proposition: Let p and q be points in $C(A^+)$ and $C(A^-)$ respectively that minimize the distance between these two sets. Let $w = p - q$, $B^+ = w \cdot p$ and $B^- = w \cdot q$.



Define hyperplanes

$$f^{\pm}(x) = w \cdot x - B^{\pm} = 0.$$

Then $f^{\pm} = 0$ are supporting hyperplanes for $C(A^{\pm})$ respectively and the associated margin

$$\tau_w(A^+, A^-) = \frac{B^+ - B^-}{\|w\|} = \|p - q\| = \frac{w \cdot (p - q)}{\|w\|}$$

$$= \frac{\|w\|^2}{\|w\|} = \|w\| = \|p - q\|$$

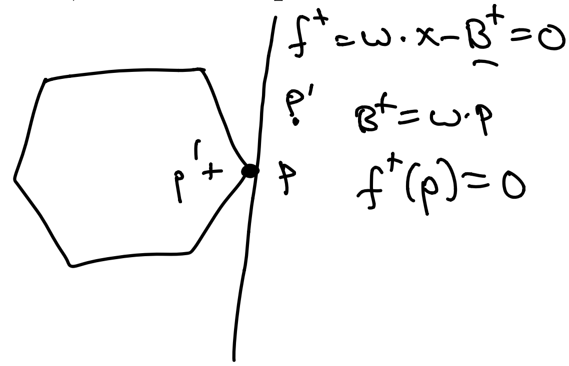
is optimal.

Proof of closest points yield optimal margin

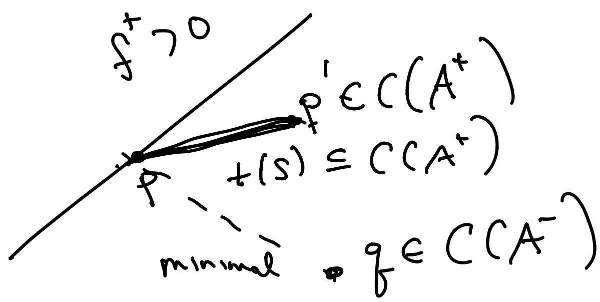
First Part: $f^{\pm}(x) = 0$ are supporting hyperplanes.

Consider f^+. If it is not a supporting hyperplane, there is a point $p' \in C(A^+)$ so that $f(p') < 0$.

$\underline{f(p') < 0}$

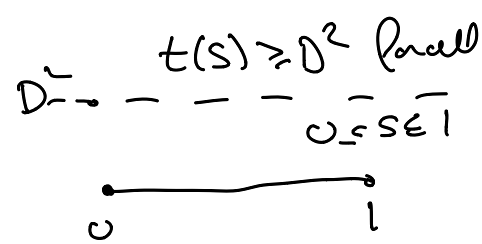


Look at the line segment $t(s) = (1 - s)p + sp'$ joining p to p', which lies inside $C(A^+)$.



Consider the distance $D(s) = \|t(s) - q\|^2$ from points on this line segment to q.

$$D(s) = \|t(s) - q\|^2$$
$$D(0) = \|p - q\|^2 = D^2$$
$$D(1) = \|p' - q\|^2 \geq D^2$$

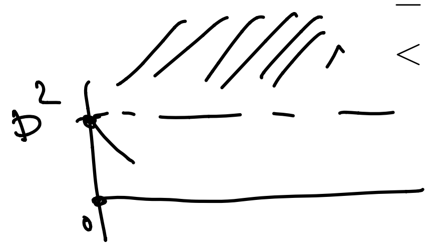


$$\|t(s)-q\|^2$$

$$2(t(s)-q)(\dot{t}(s)) = 2(t(s)-q)\cdot(p'-p)$$

We have

$$
\begin{aligned}
\frac{dD(s)}{ds}|_{s=0} &= 2(p-q)\cdot(p'-p) = 2w\cdot(p'-p)\\
&= 2((f^+(p')+B^+)-(f^+(p)+B^+))\\
&= 2f^+(p') < 0\\
&< 0
\end{aligned}
$$



$$w\cdot p'$$

$$f^+(p') = w\cdot p' - B^+$$

$$f^+(p) = 0$$

As in earlier arguments, this is impossible since p is the closest point to q in $C(A^+)$.

Similarly, f^- is a supporting hyperplane.

Finishing the proof

Second Part: The margin is $\|p - q\|$.

Remember that $w = p - q$. Then

$$\tau_w = \frac{B^+ - B^-}{\|w\|} = \frac{w \cdot (p - q)}{\|w\|} = \|p - q\|$$

This is as large as possible, so it is optimal.