2/12/2020.

Linear Classifiers:- specified by a hyperplane $(\bar{w}, t)$ in $\mathbb{R}^d$, and for a feature $\bar{x} \in \mathbb{R}^d$,

predicted label $\hat{y} = \text{sign}(\bar{w} \cdot \bar{x} - t)$.

($\bar{w} \in \mathbb{R}^d$, $t \in \mathbb{R}$)

Perceptron:- An algo for linear classifiers.

① $\bar{w} = \bar{0}, t = 0$

② REPEAT:-

for $i = 1, \ldots, n$ ← $\approx (n \times 4/\gamma^2)$

if $\text{sign}(\bar{w}\bar{x}_i - t) \neq y_i$,

$\bar{w} \leftarrow \bar{w} + y_i \bar{x}_i$

$t \leftarrow t - y_i$

Theorem. The perceptron algorithm learns a perfect classifier of a **separable** dataset with margin γ, and $\|\bar{x}_i\| \leq 1$ with at most $4/\gamma^2$ updates.

$(\bar{w}_{opt}, t_{opt})$ optimal classifier w/ margin γ, $\|\bar{w}_{opt}\| = 1$.

① $\|(w_{opt}, t_{opt})\| \leq \sqrt{2}$

② $(\bar{w}_j, t_j) \cdot (\bar{w}_{opt}, t_{opt}) \geq j\gamma$

③ $\|\bar{w}_j, t_j\| \leq \sqrt{2j}$

} we proved these last time.

For any vectors $\bar{u}, \bar{v}$, $\bar{u} \cdot \bar{v} \leq \|\bar{u}\| \cdot \|\bar{v}\|$

since cosine(angle between $\bar{u}, \bar{v}$) $= \dfrac{\bar{u} \cdot \bar{v}}{\|\bar{u}\| \|\bar{v}\|} \leq 1$

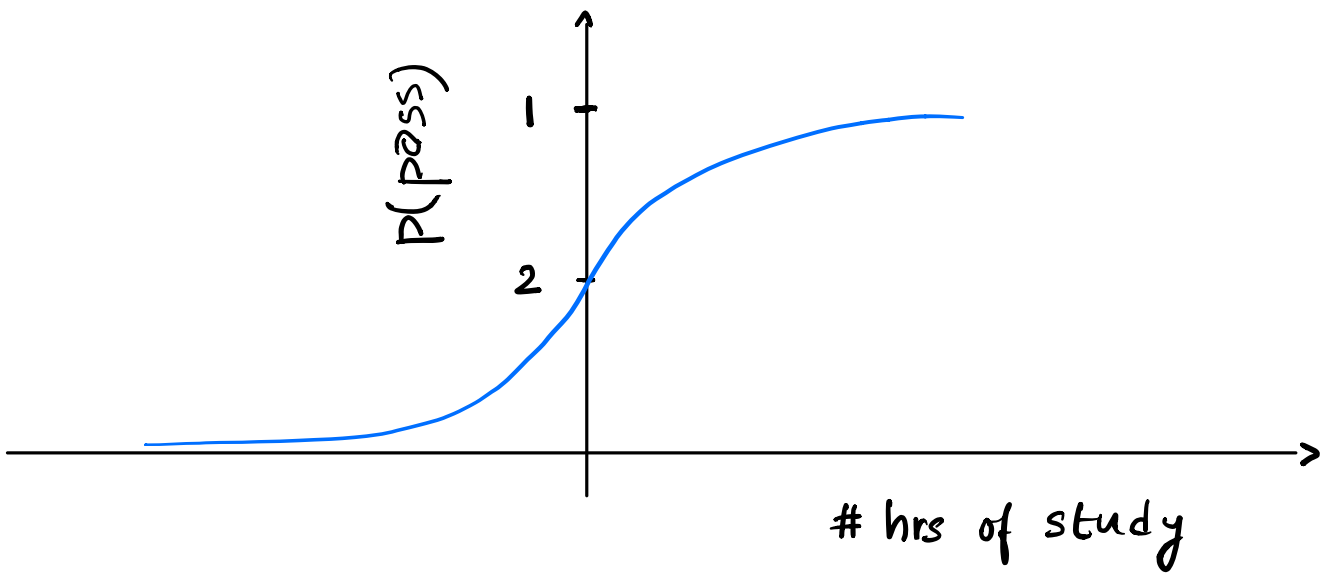
$\Rightarrow \frac{j\gamma}{\sqrt{2j}\cdot\sqrt{2}} \leq 1 \Rightarrow \boxed{j \leq \frac{4}{\gamma^2}}$

$\frac{(\bar{w}_j, t_j)\cdot(w_{opt}, t_{opt})}{\|\bar{w}_j, t_j\| \; \|w_{opt}, t_{opt}\|}$

LOGISTIC REGRESSION.

Logistic models:- a powerful model for the probability of success (labels) given the variables (features)

EX:- What is the probability of passing an exam if one studies for 'x' hours?

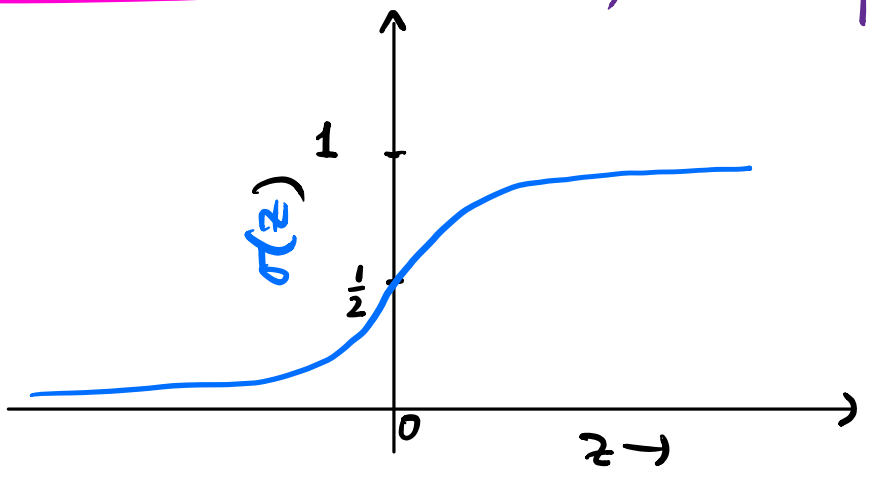


Model:- $effort = (a \cdot \#\text{ hours studied} + b)$

$$\sigma(effort) = Prob(pass) = \frac{e^{effort}}{1 + e^{effort}} = \frac{\exp(effort)}{1 + \exp(effort)}.$$

What is the plot above? it is the sigmoid function.

SIGMOID FUNCTION:- For $z \in \mathbb{R}$, $\sigma(z) := \frac{e^z}{1+e^z}$.



$\bar{x} = (\bar{x}^1, \ldots, \bar{x}^d)$ $\qquad y \in \{-1, 1\}$.

$(\bar{w}, t) \longrightarrow ((\bar{w}^1, \ldots, \bar{w}^d), t)$. $\qquad (\bar{w}', t)$ ($\bar{w}'$: a, t: $-b$)

effort $= (\bar{w}^1 \cdot \bar{x}^1 + \ldots + \bar{w}^d \cdot \bar{x}^d - t)$

Logistic Regression:- Prob. of a class is - sigmoid of a linear combination of the attributes.

✓ $\text{Prob}\left(1 \mid \bar{x}, (\bar{w}, t)\right) = \sigma(\bar{w} \cdot \bar{x} - t) = \frac{\exp(\bar{w} \cdot \bar{x} - t)}{1 + \exp(\bar{w} \cdot \bar{x} - t)}$.

✓ $\text{Prob}\left(-1 \mid \bar{x}, (\bar{w}, t)\right) = \frac{1}{1 + \exp(\bar{w} \cdot \bar{x} - t)}$.

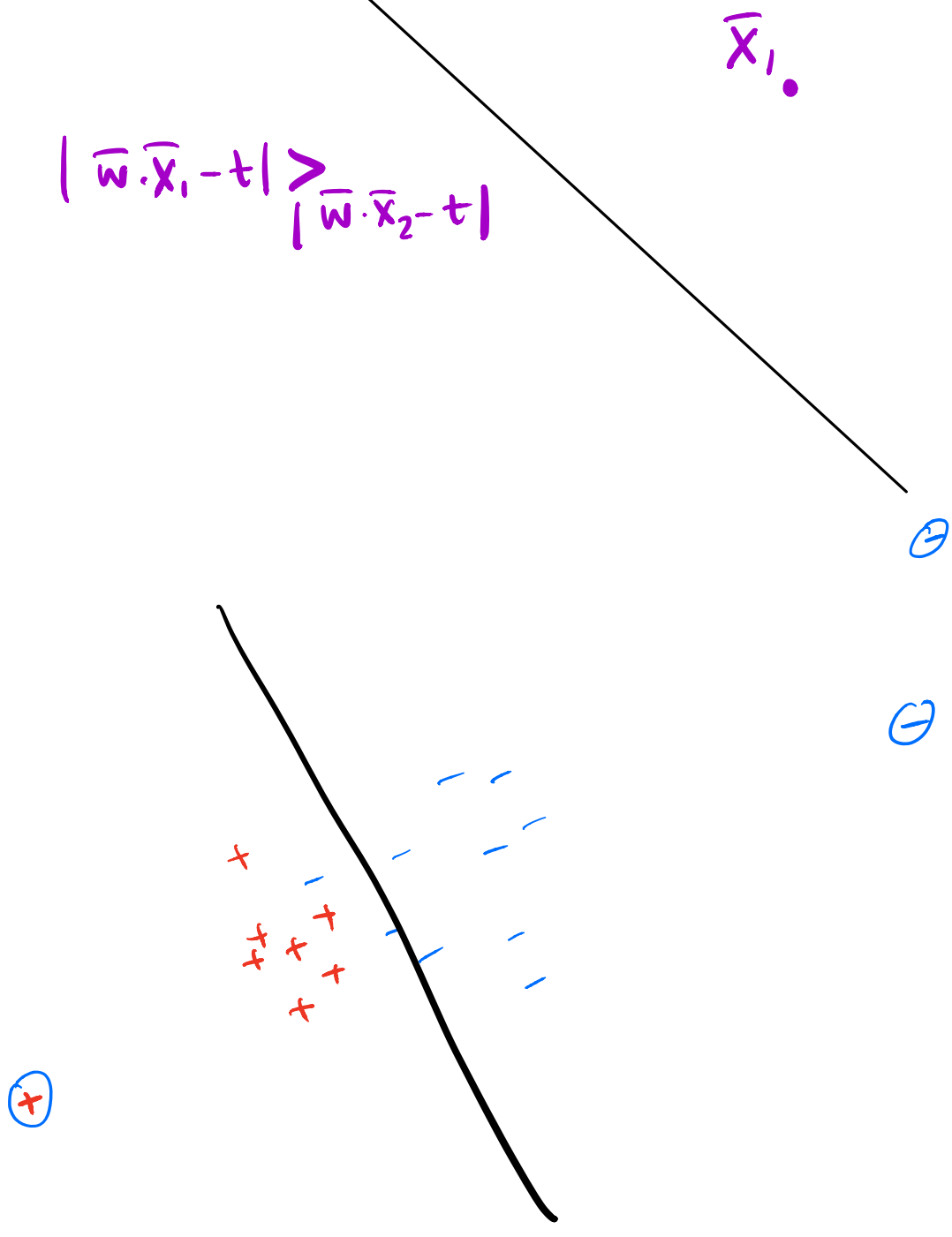
output $\max_{y \in \{-1,1\}} \{\text{Prob}(y \mid \bar{x}, (\bar{w}, t))\}$.

$\pm 1 \rightarrow \text{Prob}(1 \mid \bar{x}, (\bar{w}, t)) \geq \text{Prob}(-1 \mid \bar{x}, (\bar{w}, t))$

$\Leftrightarrow \exp(\bar{w} \cdot \bar{x} - t) \geq 1 \Leftrightarrow \bar{w} \cdot \bar{x} - t \geq 0$

$\Rightarrow \hat{y} = \text{sign}(\bar{w} \cdot \bar{x} - t)$.

$(\bar{w}, t)$ $\qquad \bar{x}_2$



$$\text{Prob}(y \mid \bar{x}, (\bar{w}, t)) = \left[\frac{\exp(\bar{w}\cdot\bar{x} - t)}{1+\exp(\bar{w}\cdot\bar{x} - t)}\right]^{\frac{1+y}{2}} \left[\frac{1}{1+\exp(\bar{w}\cdot\bar{x} - t)}\right]^{\frac{1-y}{2}}$$

$y = +1$ $\qquad$ $y = -1$

$$\text{Prob}(y \mid \bar{x}, (\bar{w}, t)) = [\exp(\bar{w}\cdot\bar{x} - t)]^{\frac{1+y}{2}} \cdot \frac{1}{1+\exp(\bar{w}\cdot\bar{x} - t)}$$

log

Assumption:- All 'n' training examples in $S = \{(\bar{x}_1, y_1) \cdots (\bar{x}_n, y_n)\}$ are generated independently.

$$\text{Prob}(y_1 \cdots y_n \mid \bar{x}_1 \cdots \bar{x}_n) = \prod_{i=1}^{n} \text{Prob}(y_i \mid \bar{x}_i).$$

$$= \prod_{i=1}^{n} \text{Prob}(y_i \mid \bar{x}_i, (\bar{w}, t))$$

$$\boxed{\max_{(\bar{w},t)} \prod_{i=1}^{n} \text{Prob}(y_i \mid \bar{x}_i, (\bar{w}, t))}.$$

- Take logs when you have product of a bunch of things.

$$\text{LLF} = \boxed{J_S(\bar{w}, t) = \sum_{i=1}^{n} \log \text{Prob}(y_i \mid \bar{x}_i, (\bar{w}, t))}$$

CLAIM:- $\max_{(\bar{w},t)} \prod_{i=1}^{n} \text{Prob}(y_i \mid \bar{x}_i, (\bar{w}, t)) = \max_{\bar{w}, t} J_S(\bar{w}, t)$

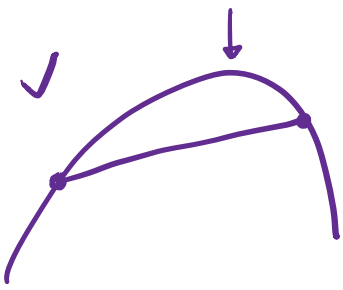
$$J_S(\bar{w}, t) = \sum_{i=1}^{n} \left[\left(\frac{1+y_i}{2} \right) \cdot (\bar{w} \cdot \bar{x}_i - t) - \log\left(1 + \exp(\bar{w} \cdot \bar{x}_i - t)\right) \right]$$

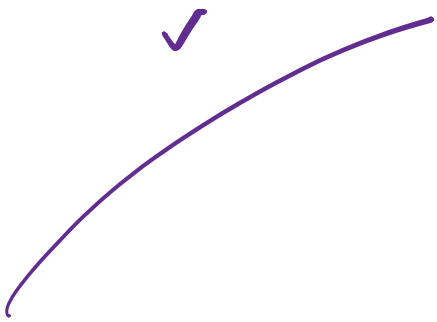
$\log(\exp(x)) = x.$

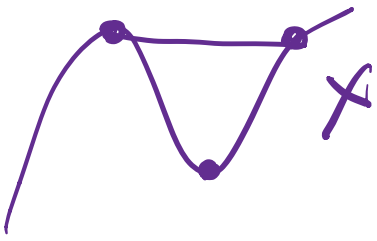
CLAIM:- $J_S(\bar{w}, t)$ is a concave function of $(\bar{w}, t)$.

$$J_S(\bar{w}_1, t_1) + J_S(\bar{w}_2, t) \leq J\left(\frac{\bar{w}_1 + \bar{w}_2}{2}, \frac{t_1 + t_2}{2}\right)$$

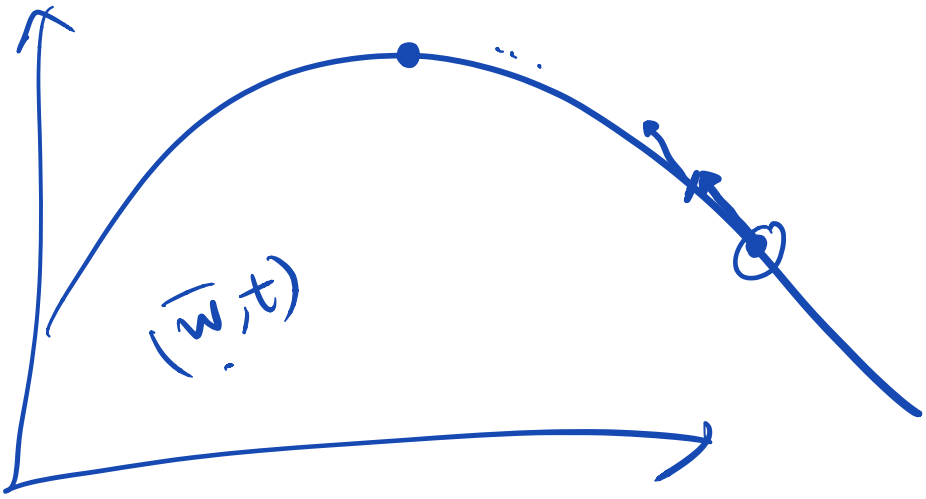
2







- **EXERCISE:-** Show $f(x) = \log(1+\exp(x))$ is convex.



LR

- Start with $(\bar{w}_0, t_0)$,
  until convergence:-

$$(\bar{w}_j, t_j) \leftarrow (\bar{w}_{j-1}, t_{j-1}) + \eta \cdot \nabla J_s(\bar{w}, t)\Big|_{(\bar{w}_{j-1}, t_{j-1})}$$

$$\nabla J_s(\bar{w}, t) = \left[\frac{\partial J(\bar{w}, t)}{\partial \bar{w}^1}, \ldots, \frac{\partial J(\bar{w}, t)}{\partial \bar{w}^d}, \frac{\partial J(\bar{w}, t)}{\partial t}\right].$$

$$J_s(\bar{w}, t) = \sum_{i=1}^{n}\left[\left(\frac{1+y_i}{2}\right)\cdot\underbrace{(\bar{w}\cdot\bar{x}_i - t)}_{\bar{w}^1\cdot\bar{x}_i^1} - \log\left(1+\exp(\bar{w}\cdot\bar{x}_i - t)\right)\right]$$

should not be there.

$$\frac{d}{dx}\log(1+\exp(x)) = \frac{1}{1+\exp(x)} \cdot \exp(x) \cdot \boxed{x} = \sigma(x) \cdot \boxed{x}$$

- $\bar{w}\cdot\bar{x} - t = (\bar{w}, t)\cdot\bar{x}^*$, $\quad \bar{x}^* = (\bar{x}^1, \ldots, \bar{x}^d, -1)$

$$\nabla J_s(\bar{w}, t) = \sum_{i=1}^{n} \left[\left(\frac{1+y_i}{2}\right) \cdot \bar{x}_i^* - \frac{\exp(\bar{w}\cdot\bar{x} - t)}{1+\exp(\bar{w}\cdot\bar{x} - t)} \cdot \bar{x}_i^* \right] \leftarrow$$

$$= \sum_{i=1}^{n} \left[\left(\frac{1+y_i}{2}\right) - P(+1 \mid \bar{x}_i^*, \bar{w}, t) \right] \bar{x}_i^*$$

EXERCISE :- derive from 1st principle (defn).