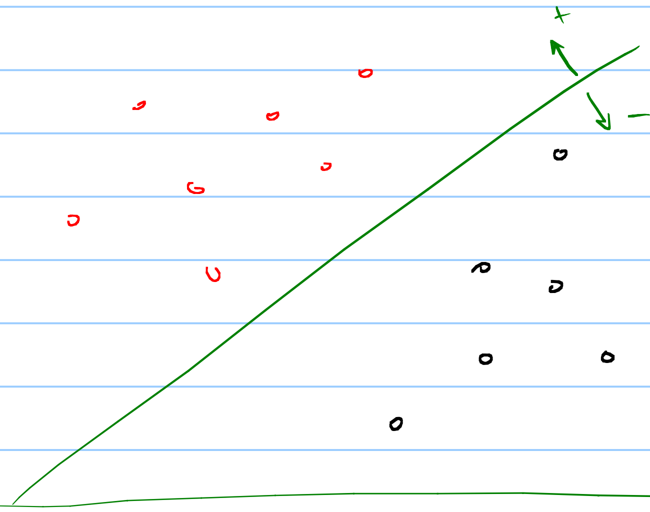
# Perceptron

$$\{(x_p, y_p)\}_{p=1}^{P}, \quad y_p \in \{-1, +1\}$$ ← labels

$y = 1$



$h(x) = b + \bar{x}^T\bar{w} = 0$

$y = -1$

$$\boxed{\begin{array}{lll} \text{if} & b + \bar{x}_p^T\bar{w} > 0 & y_p = 1 \\ \text{if} & b + \bar{x}_p^T\bar{w} < 0 & y_p = -1 \end{array}}$$

$$(b + \bar{x}_p^T\bar{w}) \cdot y_p > 0$$

$$\Rightarrow -y_p(b + \bar{x}_p^T\bar{w}) < 0$$

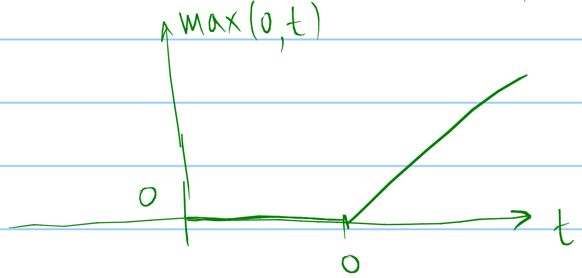
$$g = \max(0, -y_p(b + \bar{x}_p^T\bar{w})) = 0$$ (correct classification)

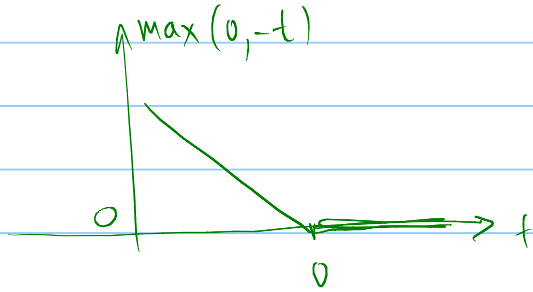
if $g > 0 \Rightarrow$ misclassification

## Loss function

$$g_1(b, \bar{w}) = \sum_{p=1}^{P} \max(0, -y_p(b + \bar{x}_p^T\bar{w}))$$

perceptron, hinge cost, max cost function, rectified linear unit



$\max(0,-t)$ → convex
→ non-differentiable at 0
→ trivial solution $t=0$

$\max(0,-y_p(b+\bar{x}_p^T\bar{w}))$ : $g_1=0$ for $b=0, \bar{w}=0$

Approximations of $\max(0,t)$ function.

(I) softmax function : $\text{soft}(s_1,s_2)=\log(e^{s_1}+e^{s_2})$

we can show that $\text{soft}(s_1,s_2)\simeq \max(s_1,s_2)$

$\therefore\ s_1 >> s_2 \longrightarrow e^{s_1} >> e^{s_2} \longrightarrow \log(e^{s_1}+e^{s_2}) \sim \log(e^{s_1}) = s_1$

$\hookrightarrow \max(s_1,s_2)=s_1$

$s_1 << s_2 \longrightarrow e^{s_2} >> e^{s_1} \longrightarrow \log(e^{s_1}+e^{s_2}) \sim \log(e^{s_2}) = s_2$

$$g_1=\max(0,-y_p(b+\bar{x}_p^T\bar{w})) \approx \log(e^0+e^{-y_p(b+\bar{x}_p^T\bar{w})})$$
$$=\log(1+e^{-y_p(b+\bar{x}_p^T\bar{w})})=g_2$$

{ logistic regression
softmax perceptron
log-loss SVM }

1) differentiable

2) $b=0,\ \bar{w}=0 \rightarrow g_2=\log(1+1)=\log(2)>0$

---

$$g_2(b,\bar{w})=\sum_{p=1}^{P}\log(1+e^{-y_p(b+\bar{x}_p^T\bar{w})})$$

$$\tilde{x}_p=\begin{bmatrix}1\\ \bar{x}_p\end{bmatrix},\quad \tilde{w}=\begin{bmatrix}b\\ \bar{w}\end{bmatrix}$$

$$g_2(\tilde{w}) = \sum_{p=1}^{P} \log\left(1 + e^{-y_p \tilde{x}_p^T \tilde{w}}\right)$$

## Gradient

$$-y_p \tilde{x}_{p1} \cdot w_1$$

$$\frac{\partial g_2}{\partial w_1} = \sum_{p=1}^{P} \frac{1}{1+e^{-y_p \tilde{x}_p^T \tilde{w}}} \cdot e^{-y_p \tilde{x}_p^T \tilde{w}} \cdot (-y_p \tilde{x}_{p1})$$

$$= \frac{1}{e^{+y_p \tilde{x}_p^T \tilde{w}}\left(1+e^{-y_p \tilde{x}_p^T \tilde{w}}\right)} = \frac{1}{1+e^{y_p \tilde{x}_p^T \tilde{w}}}$$

sigmoid: $\sigma(t) = \dfrac{1}{1+e^{-t}} \longrightarrow \sigma(-t) = \dfrac{1}{1+e^{t}}$

$$\nabla_{\tilde{w}} g_2 = \begin{bmatrix} \frac{\partial g_2}{\partial w_1} \\ \frac{\partial g_2}{\partial w_2} \\ \vdots \\ \frac{\partial g_2}{\partial w_{N+1}} \end{bmatrix} = \begin{bmatrix} -\sum_{p=1}^{P} \sigma(-y_p \tilde{x}_p^T \tilde{w}) \cdot y_p \tilde{x}_{p1} \\ -\sum_{p=1}^{P} \sigma(-y_p \tilde{x}_p^T \tilde{w}) \cdot y_p \tilde{x}_{p2} \\ \vdots \\ -\sum_{p=1}^{P} \sigma(-y_p \tilde{x}_p^T \tilde{w}) \cdot y_1 \cdot \tilde{x}_{p(N+1)} \end{bmatrix} = -\sum_{p=1}^{P} \sigma(-y_p \tilde{x}_p^T \tilde{w}) y_p \tilde{x}_p$$

## Hessian matrix

$$\frac{d\sigma(t)}{dt} = \sigma(t)\left[1-\sigma(t)\right]$$

find $\dfrac{\partial g_2}{\partial w_1 \partial w_2} = \dfrac{\partial}{\partial w_2}\left[-\sum_p \sigma(\underbrace{-y_p \tilde{x}_p \tilde{w}}_{a}) y_p \tilde{x}_{p1}\right]$

$$= -\sum_p \sigma(a)(1-\sigma(a))(-y_p \tilde{x}_{p2}) \cdot y_p \tilde{x}_{p1}$$

$$= \sum_p \sigma(a)(1-\sigma(a)) \tilde{x}_{p1} \tilde{x}_{p2}$$

$$\nabla^2 g_2 = \sum_p \sigma(-y_p \tilde{x}_p^T \tilde{w})\left[1-\sigma(-y_p \tilde{x}_p^T \tilde{w})\right] \underbrace{\tilde{x}_p \cdot \tilde{x}_p^T}_{\text{matrix}}$$