convex = chord lies above function evaluated between (W_0, W_1)

Find W* $\dfrac{\partial L(w)}{\partial w} = 0$ $= \sum_i -y_i \mathbf{x}_i (1 - p(y_i|\mathbf{x}_i))$

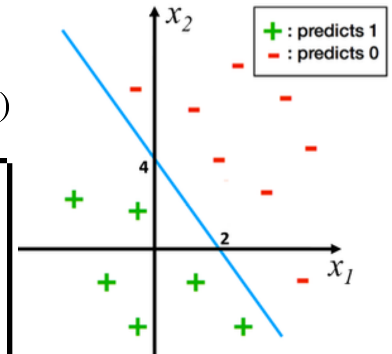
$$h(x_1, x_2) = \begin{cases} 1, if\ w_1x_1 + w_2x_2 + 8 \geq 0 \\ 0, otherwise \end{cases}$$

*first, find points where y = 0 & x = 0

*solve for (W_0, W_1)

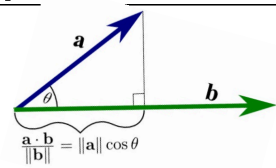
$$w_1x_1 + w_2x_2 + 8 = 0$$
$$w_1(2) + w_2(0) + 8 = 0$$
$$2w_1 + 0 + 8 = 0$$
$$2w_1 = -8$$
$$w_1 = -4$$

$$w_1x_1 + w_2x_2 + 8 = 0$$
$$w_1(0) + w_2(4) + 8 = 0$$
$$0 + 4w_2 + 8 = 0$$
$$4w_2 = -8$$
$$w_2 = -2$$



B. Linear regression deals with the prediction of continuous values; logistic regression deals with the prediction of class labels.

$$AB = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$
$$= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$$



Dot Product: $a \cdot b = a^{\intercal} b$

-dot product will change value if one of the vectors gets longer but stays pointed in the same direction

$$\frac{\text{sparrow}\cdot\text{chipmunk}}{||\text{sparrow}||_2 \times ||\text{chipmunk}||_2} = 0.0082$$

-unscaled numericals in vector

$$w^* = \arg\max_w G(w)$$
$$= \arg\max_w \ln(G(w))$$
$$= \arg\min_w -\ln(G(w))$$

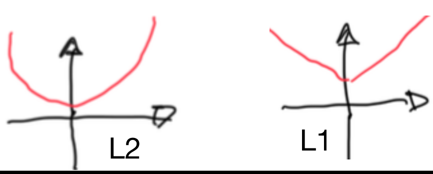
a is not a function of **x**	$\frac{\partial(\mathbf{a}\cdot\mathbf{x})}{\partial\mathbf{x}} = \frac{\partial(\mathbf{x}\cdot\mathbf{a})}{\partial\mathbf{x}} =$ $\frac{\partial\mathbf{a}^\top\mathbf{x}}{\partial\mathbf{x}} = \frac{\partial\mathbf{x}^\top\mathbf{a}}{\partial\mathbf{x}} =$	$\mathbf{a}^\top$	$\mathbf{a}$
A is not a function of **x**; **b** is not a function of **x**	$\frac{\partial\mathbf{b}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}} =$	$\mathbf{b}^\top\mathbf{A}$	$\mathbf{A}^\top\mathbf{b}$
A is not a function of **x**	$\frac{\partial\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}} =$	$\mathbf{x}^\top(\mathbf{A}+\mathbf{A}^\top)$	$(\mathbf{A}+\mathbf{A}^\top)\mathbf{x}$
A is not a function of **x**; **A** is symmetric	$\frac{\partial\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}} =$	$2\mathbf{x}^\top\mathbf{A}$	$2\mathbf{A}\mathbf{x}$
A is not a function of **x**	$\frac{\partial^2\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}^2} =$	$\mathbf{A}+\mathbf{A}^\top$	
A is not a function of **x**; **A** is symmetric	$\frac{\partial^2\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}^2} =$	$2\mathbf{A}$	

OLS Closed Form

$$W^* = (X^TX)^{-1}X^TY$$

```
A = np.dot(np.transpose(X), X)
# Obtain optimal W
W = np.dot(inv(A),np.dot(np.transpose(X),Y))
```

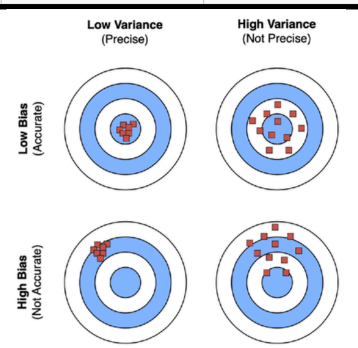
Training set -> set the parameters
Validation set -> try different models, select best
Test set -> how good is your chosen model

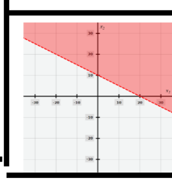


$\epsilon_{\text{testing}} = \epsilon_{\text{training}} + \epsilon_{\text{generalization}}$

train >0, gen = 0
completely random
highly generalizable, no error

train =0, gen > 0
perfect classifications
non-generalizable, large error



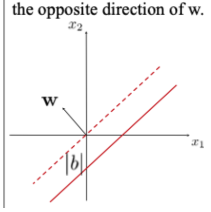


Logistic classifier decision boundary

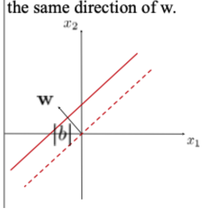
Identities: vector-by-vector $\frac{\partial\mathbf{y}}{\partial\mathbf{x}}$

Condition	Expression	Numerator layout, i.e. by **y** and $\mathbf{x}^\top$	Denominator layout, i.e. by $\mathbf{y}^\top$ and **x**
a is not a function of **x**	$\frac{\partial\mathbf{a}}{\partial\mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial\mathbf{x}}{\partial\mathbf{x}} =$	$\mathbf{I}$	
A is not a function of **x**	$\frac{\partial\mathbf{A}\mathbf{x}}{\partial\mathbf{x}} =$	$\mathbf{A}$	$\mathbf{A}^\top$
A is not a function of **x**	$\frac{\partial\mathbf{x}^\top\mathbf{A}}{\partial\mathbf{x}} =$	$\mathbf{A}^\top$	$\mathbf{A}$
a is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial a\mathbf{u}}{\partial\mathbf{x}} =$	$a\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$	
$a = a(\mathbf{x})$, **u** = **u**(**x**)	$\frac{\partial a\mathbf{u}}{\partial\mathbf{x}} =$	$a\frac{\partial\mathbf{u}}{\partial\mathbf{x}} + \mathbf{u}\frac{\partial a}{\partial\mathbf{x}}$	$a\frac{\partial\mathbf{u}}{\partial\mathbf{x}} + \frac{\partial a}{\partial\mathbf{x}}\mathbf{u}^\top$
A is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial\mathbf{A}\mathbf{u}}{\partial\mathbf{x}} =$	$\mathbf{A}\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$	$\frac{\partial\mathbf{u}}{\partial\mathbf{x}}\mathbf{A}^\top$
u = **u**(**x**), **v** = **v**(**x**)	$\frac{\partial(\mathbf{u}+\mathbf{v})}{\partial\mathbf{x}} =$	$\frac{\partial\mathbf{u}}{\partial\mathbf{x}} + \frac{\partial\mathbf{v}}{\partial\mathbf{x}}$	
u = **u**(**x**)	$\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{x}} =$	$\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$	$\frac{\partial\mathbf{u}}{\partial\mathbf{x}}\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}}$

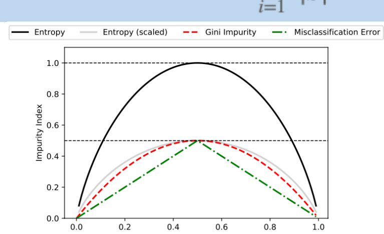
When b>0 the decision boundary is moved along the opposite direction of w.



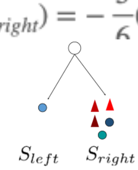
When b<0 the decision boundary is moved along the same direction of w.



$$\arg\max_B G(S,B) = H(S) - \sum_{i=1}^{t} \frac{|S_i|}{|S|} H(S_i)$$



$$G(S_{right}) = -\frac{5}{6}(0.4\log_2 0.4 + 0.6\log_2 0.6) = 0.81,$$
$$\therefore G(S,B) = H(S) - 0.81$$



$$G(S_{left}) = -\frac{1}{6}(1\log_2 1 + 0\log_2 0) = 0,$$