ECS 171: Machine Learning

Summer 2023

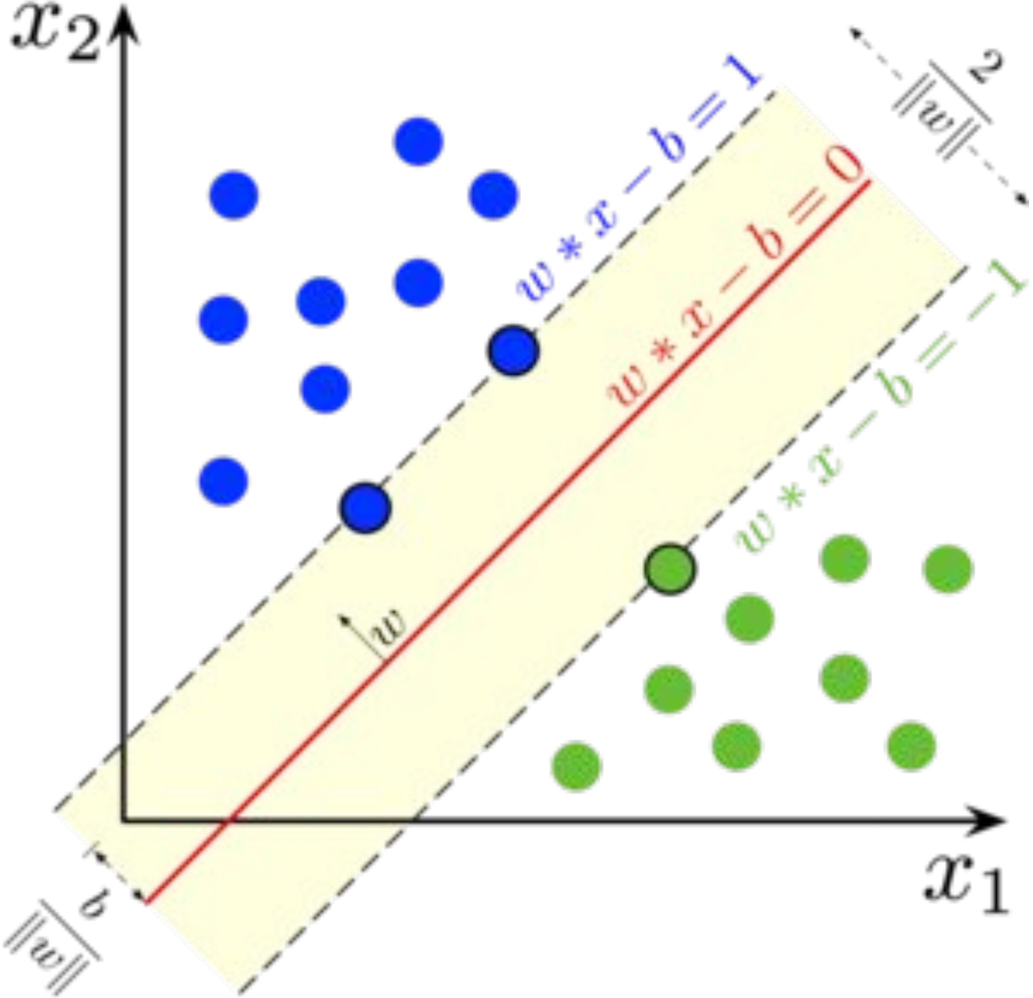
Edwin Solares
easolares@ucdavis.edu

Support Vector Machines

& Logistic Regression Again!?

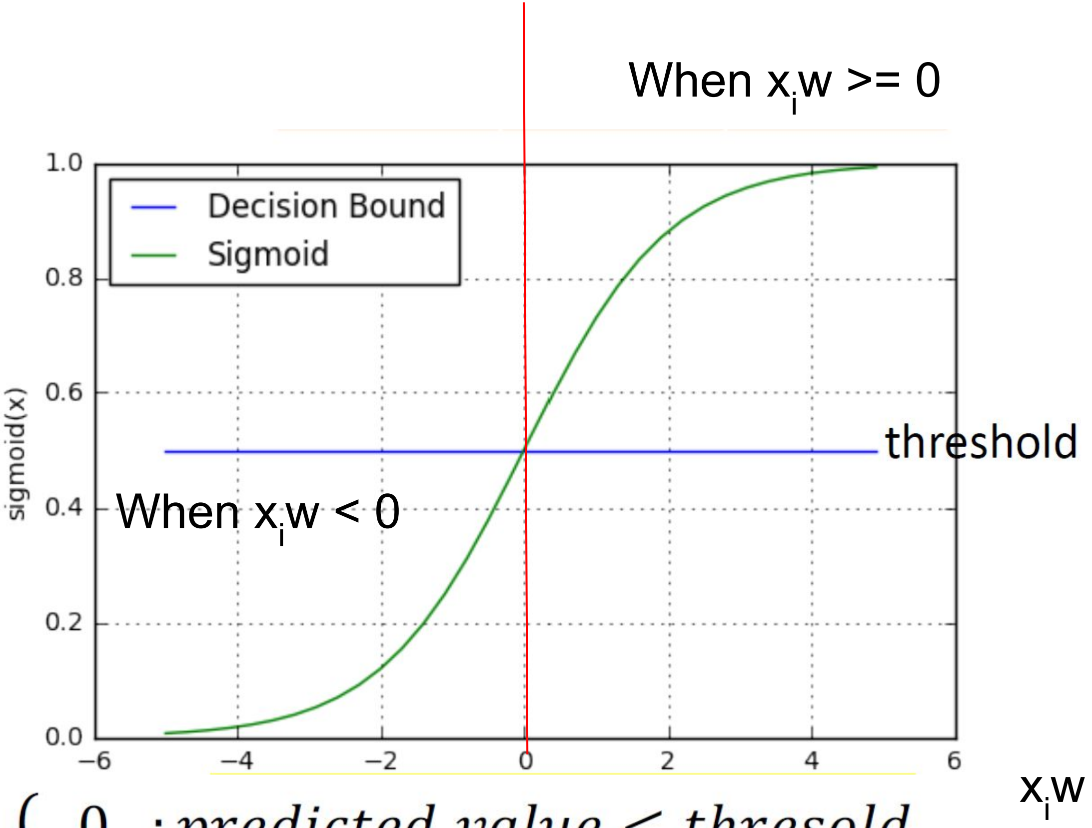
SVM Classification

Wikipedia



Logistic Regression

$\hat{y}_i$ Is our prediction given formula on the right, but is given a value of 0 or 1 based on the threshold of 0.5
Prior to threshold we can say our raw value is our probability. As our raw value is continuous from 0 to 1.



$$\hat{y}_i = \begin{cases} 0 & ; predicted\ value < thresold \\ 1 & ; predicted\ value \geq threshold \end{cases}$$

Logistic Regression

Where we have Bernoulli observations
And p_k is the probability of y_k=1 and
$1-p_k$ is the probability $y_k = 0$

The log loss for the *k*-th point is:

We can say p_k is our raw value, which is our probability of of y_k = 1 given $x_i w$

$$\text{Cost} \begin{cases} -\ln p_k & \text{if } y_k = 1, \\ -\ln(1-p_k) & \text{if } y_k = 0. \end{cases}$$

$$-y_k \ln p_k - (1-y_k)\ln(1-p_k)$$

log-likelihood

$$\ell = \sum_{k:y_k=1} \ln(p_k) + \sum_{k:y_k=0} \ln(1-p_k) = \boxed{\sum_{k=1}^{K} (y_k \ln(p_k) + (1-y_k)\ln(1-p_k))}$$

$J(w)$ = d/dw ℓ* 1/k

Logistic Regression

log-likelihood

$$\ell = \sum_{k:y_k=1} \ln(p_k) + \sum_{k:y_k=0} \ln(1-p_k) = \sum_{k=1}^{K} (\, y_k \ln(p_k) + (1-y_k)\ln(1-p_k))$$

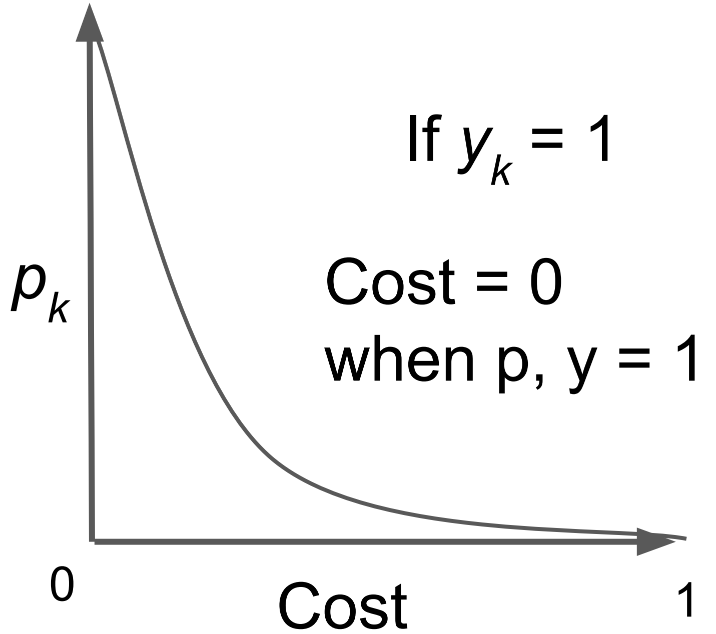
J(w) = 1/k d/dw ℓ

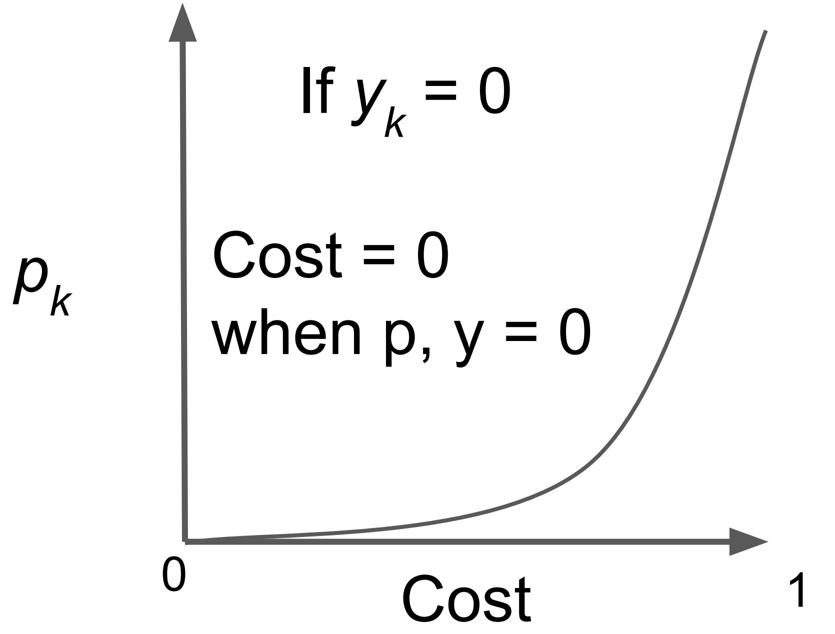
d/dw ℓ = $\sum_{k=1}^{K} (\, y_k - p_k) + \, (y_k - p_k) x_k$

Logistic Regression

The log loss for the k-th point is:

$$\text{Cost}(p_k, y_k) \begin{cases} -\ln p_k & \text{if } y_k = 1, \\ -\ln(1 - p_k) & \text{if } y_k = 0. \end{cases}$$





Logistic Regression

$$\hat{y} = \frac{1}{1 + e^{(\mu - x)/s}}$$

$\hat{y}_i$ Is our prediction given formula on the right, but is given a value of 0 or 1 based on the threshold of 0.5

Prior to threshold we can say our raw value is our probability. As our raw value is continuous from 0 to 1.

$$\hat{y} = \frac{1}{1 + e^{-(w_0 + w_1 x)}}$$

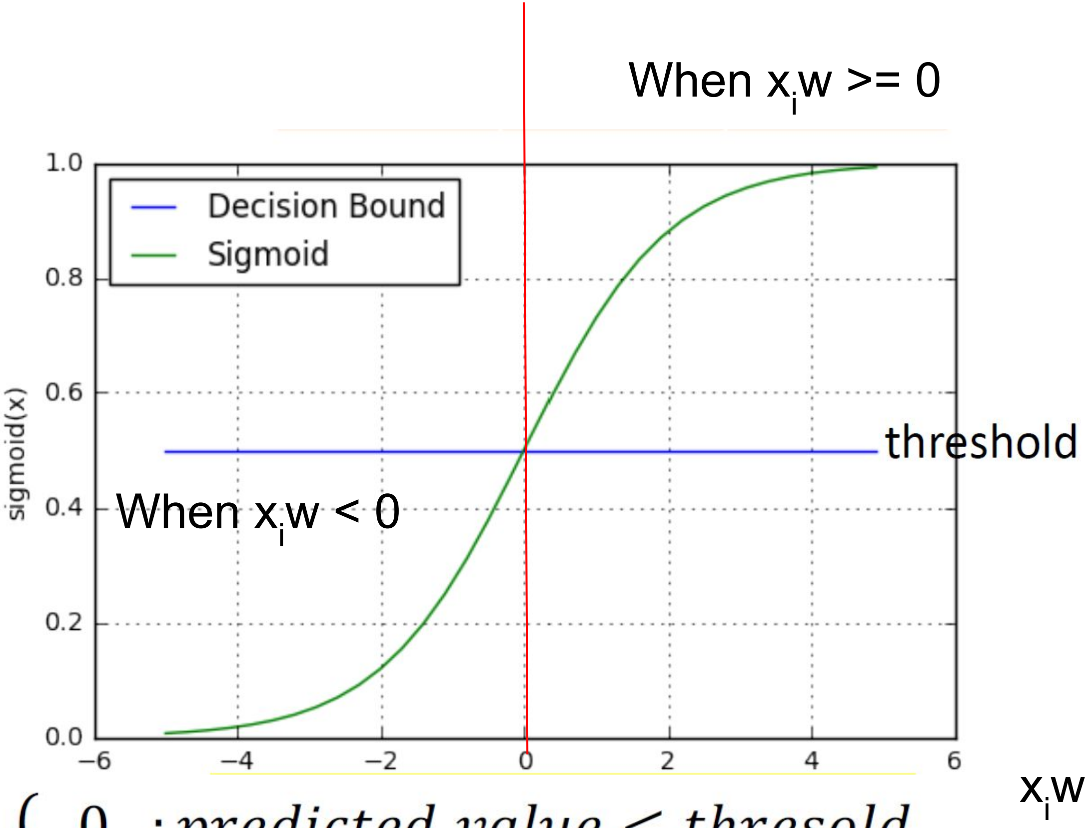
where $w_0 = -\mu/s$ and $w_1 = 1/s$ $\therefore$ we can solve for μ and s

$$\mu = w_0 / w_1 \text{ and } s = 1/w_1$$

Logistic Regression

$\hat{y}_i$ Is our prediction given formula on the right, but is given a value of 0 or 1 based on the threshold of 0.5

Prior to threshold we can say our raw value is our probability. As our raw value is continuous from 0 to 1.



$$\hat{y}_i = \begin{cases} 0 & ; predicted\ value < thresold \\ 1 & ; predicted\ value \geq threshold \end{cases}$$

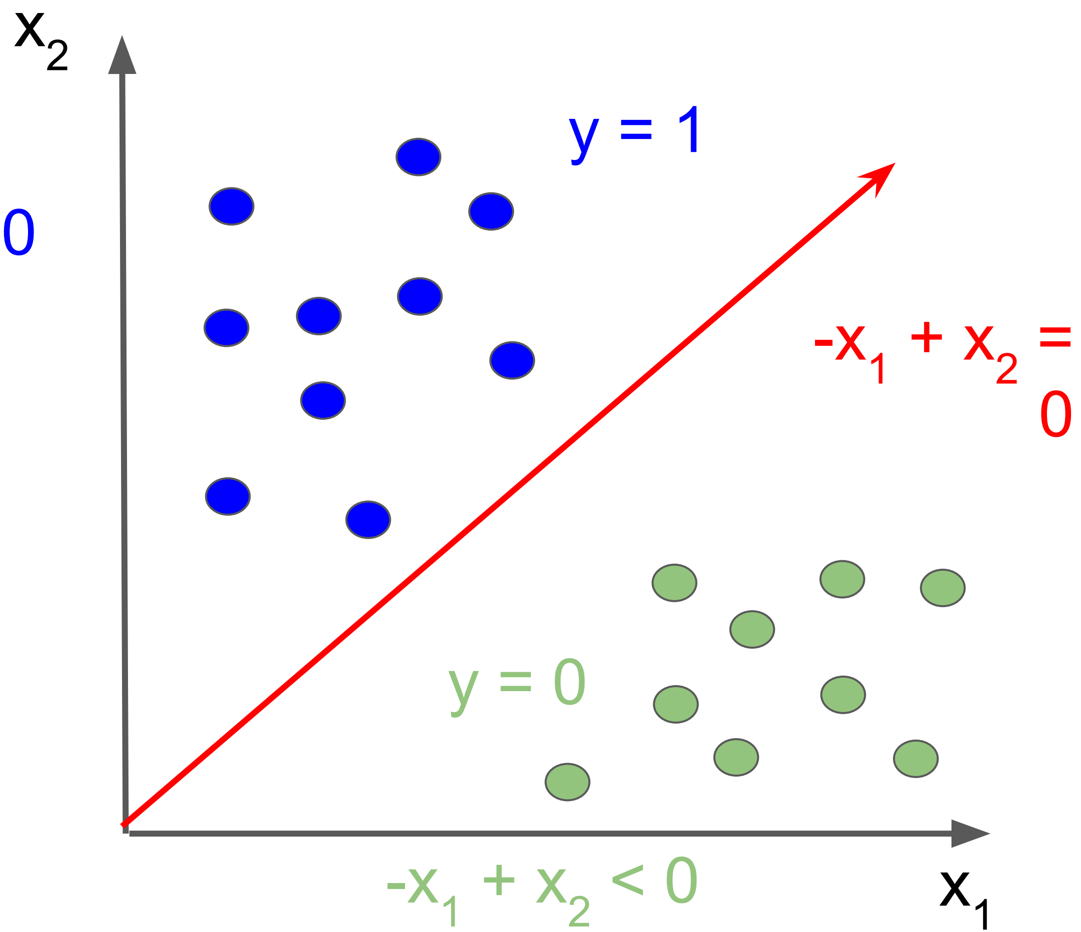
Logistic Classification

Predict $y = 1$ given $x_1 w_1 + x_2 w_2 + w_0$

$-x_1 + x_2 >= 0$

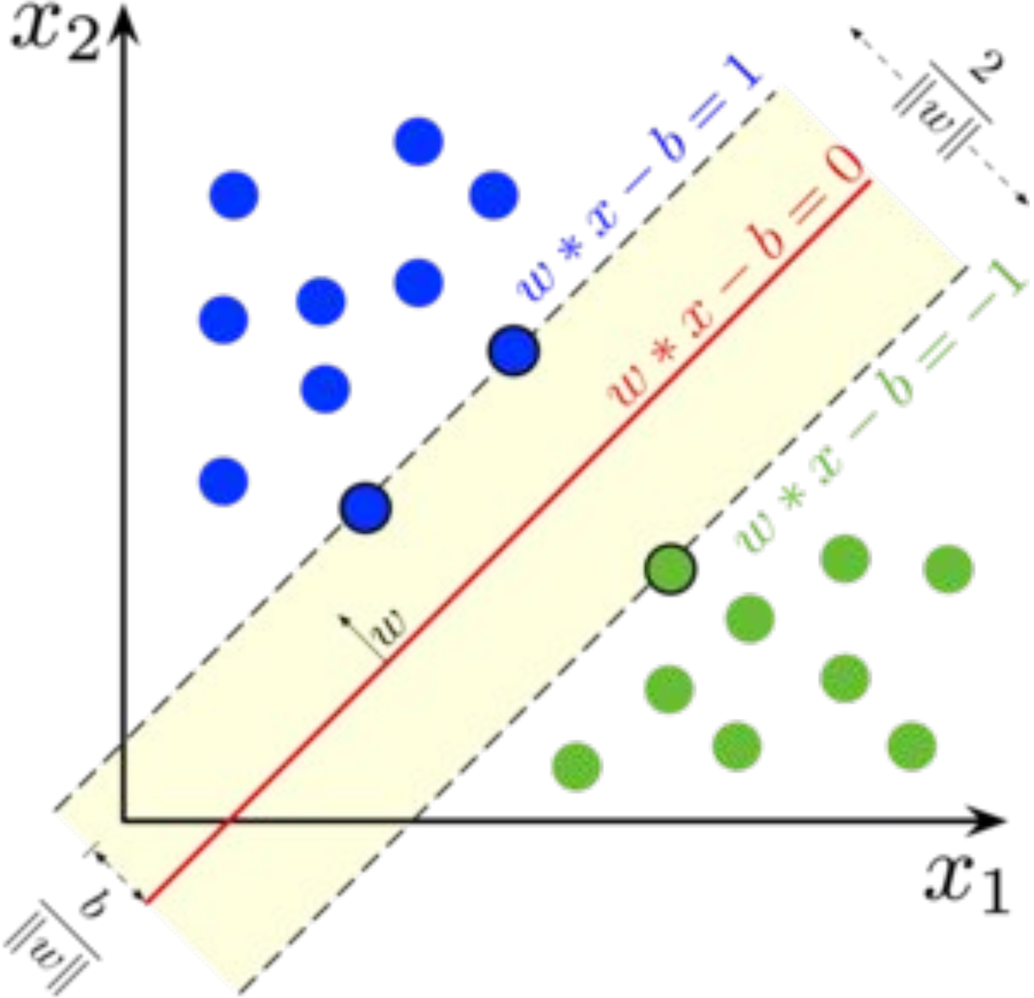
$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\mathbf{w} = \begin{bmatrix} -1 \\ +1 \end{bmatrix}$$



SVM Classification

Wikipedia



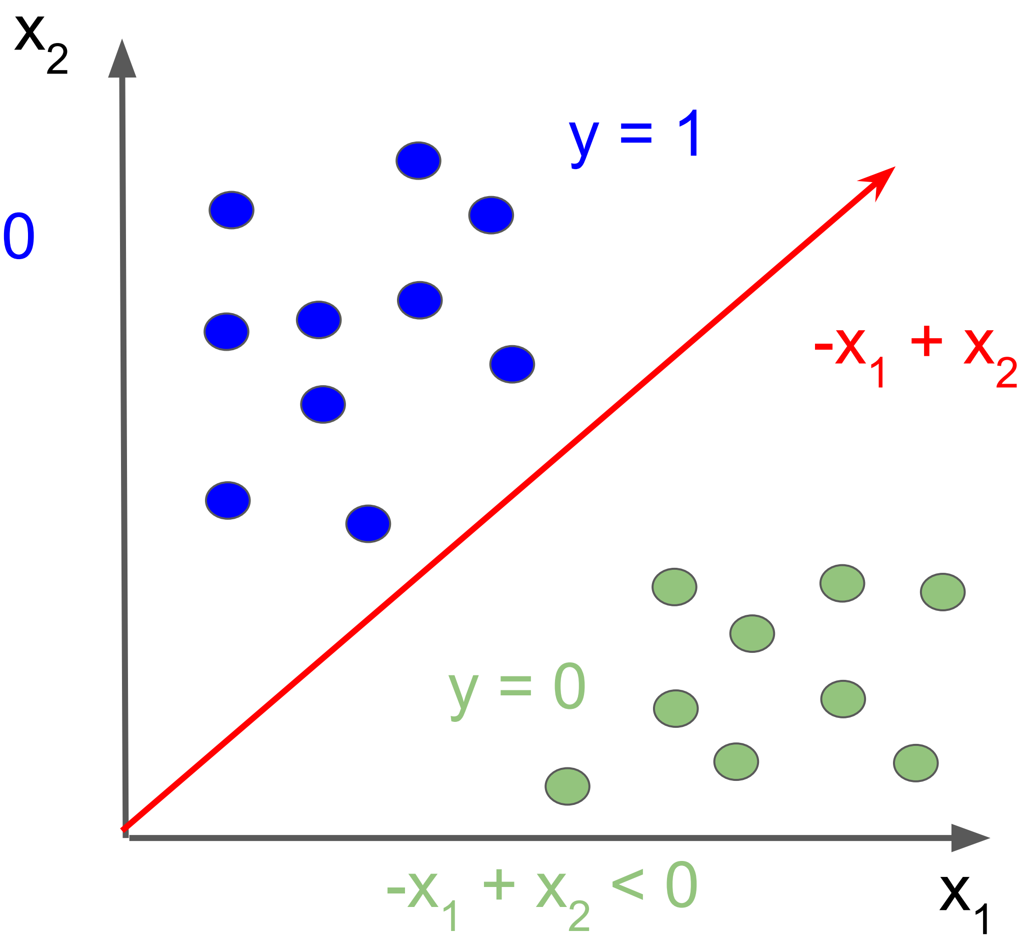
Logistic Classification

Predict $y = 1$ given $x_1w_1 + x_2w_2 + w_0$

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\mathbf{w} = \begin{bmatrix} -1 \\ +1 \end{bmatrix}$$

$-x_1 + x_2 >= 0$

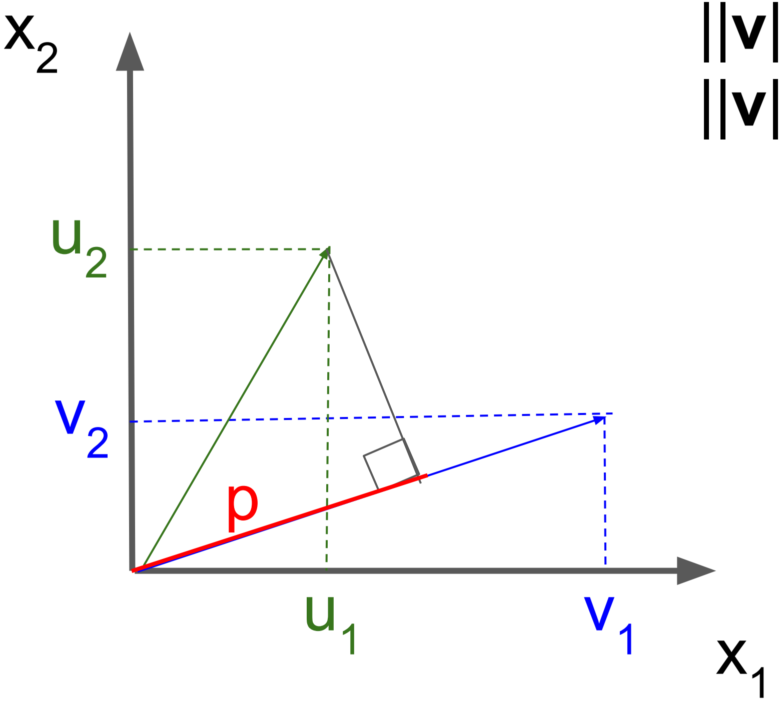


Adding Margin (Recall Inner Products)

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$||\mathbf{v}||$ = Length of $\mathbf{v}$

$$||\mathbf{v}|| = \sqrt{v_1^2 + v_2^2}$$



$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$\mathbf{v}^T\mathbf{u} = ||\mathbf{v}||\ p$

Where p is the projection of u onto v

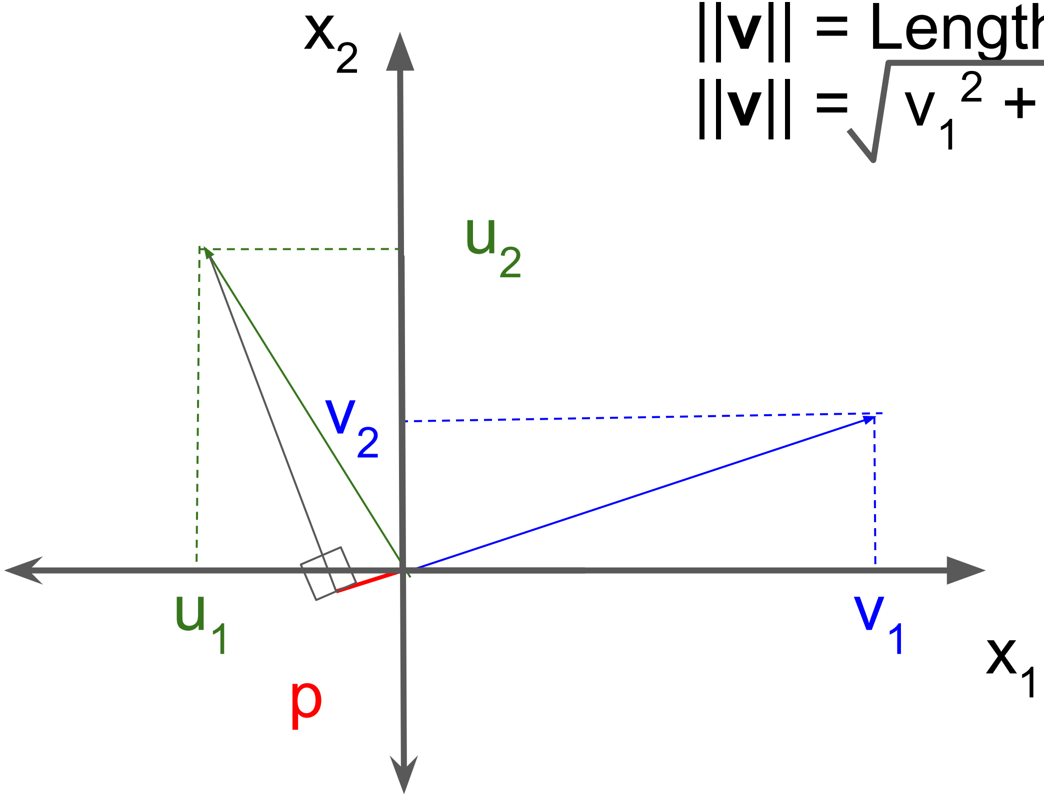
Adding Margin (Recall Inner Products)

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$||\mathbf{v}||$ = Length of $\mathbf{v}$

$$||\mathbf{v}|| = \sqrt{v_1^2 + v_2^2}$$

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$



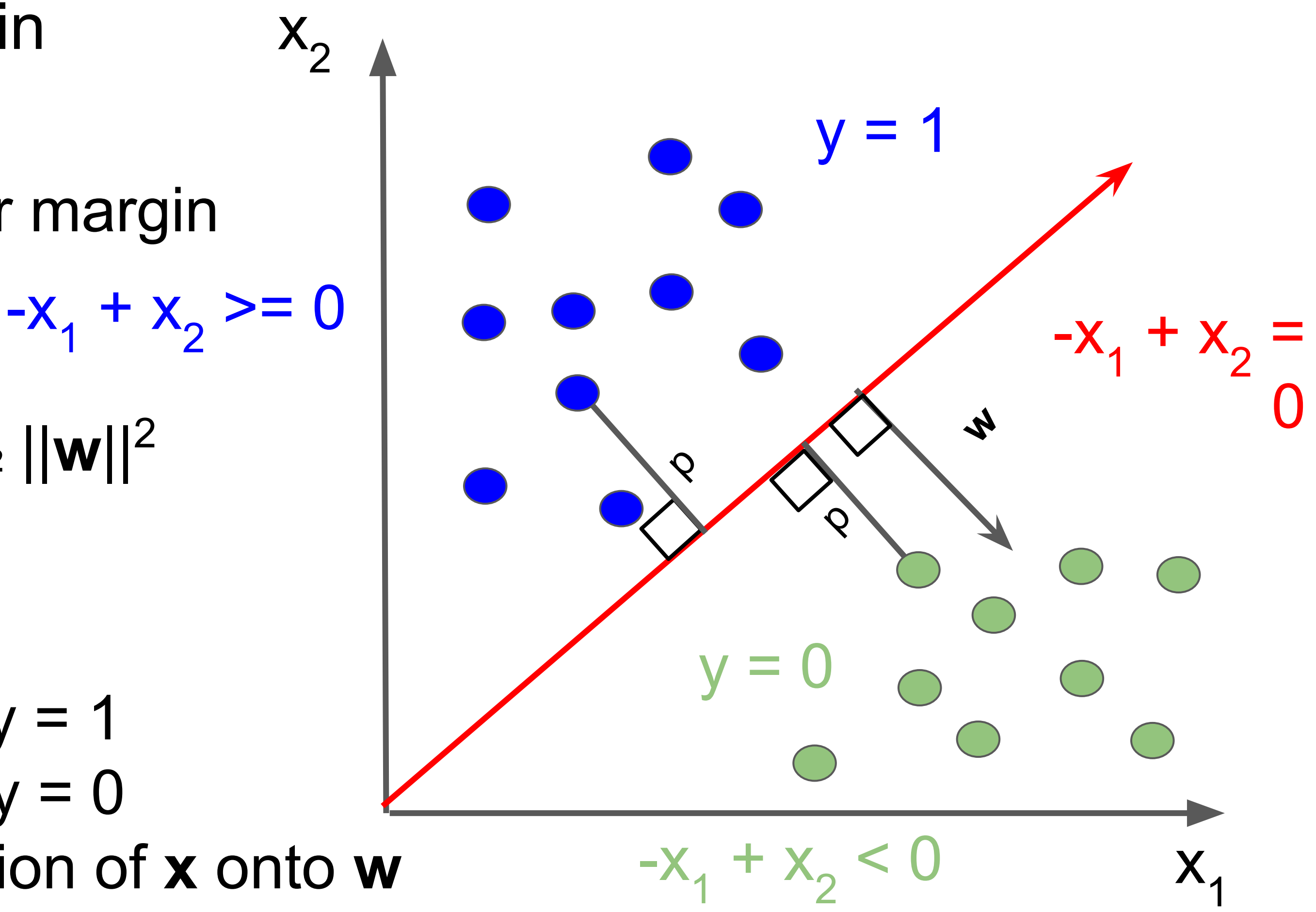
$\mathbf{v}^T\mathbf{u} = ||\mathbf{v}||\ p$

Where p is the projection of u onto v

Predict $y = 1$ given $x_1w_1 + x_2w_2 + w_0$

SVM with Margin

We want large **p**
Since it wil be our margin



$$\min_w \frac{1}{2} \sum_{j=1}^{m} w_j^2 = \frac{1}{2} ||\mathbf{w}||^2$$

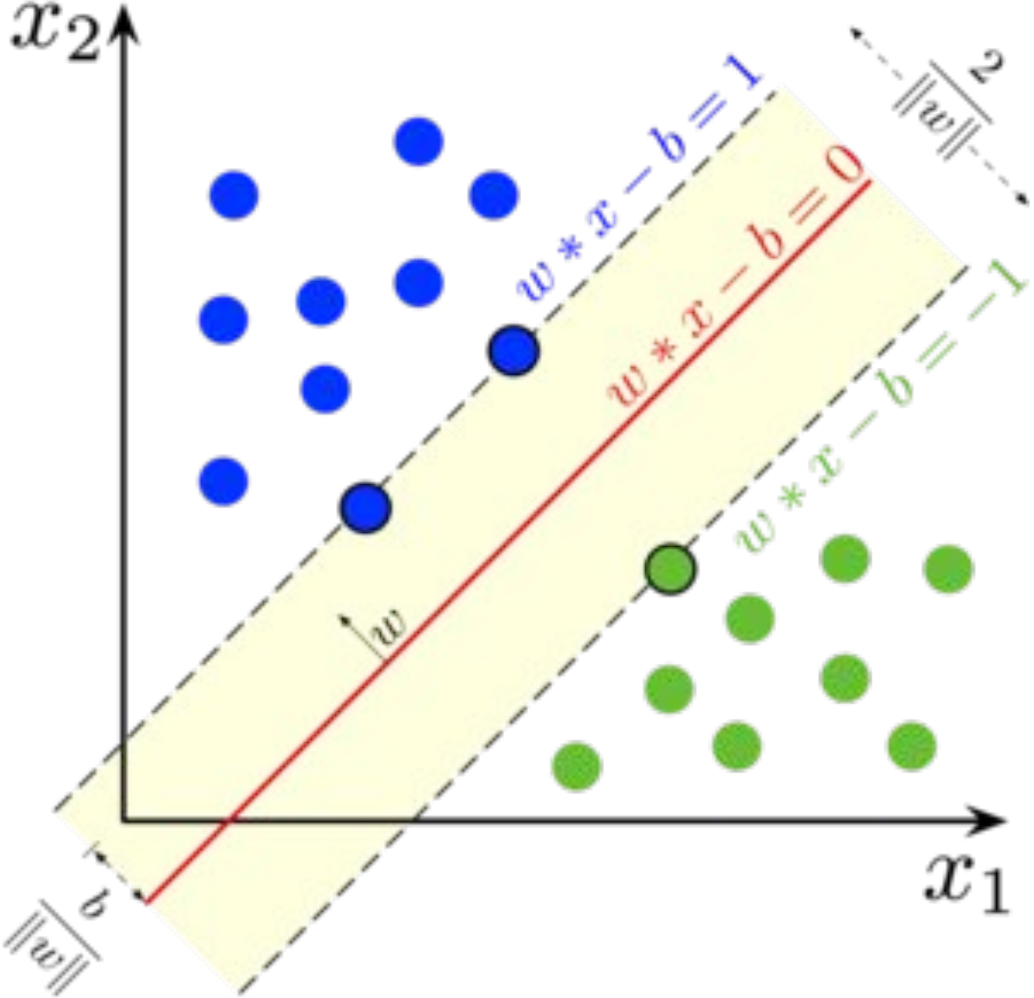
$\mathbf{p}\ ||\mathbf{w}|| => 1$ if $y = 1$

$\mathbf{p}\ ||\mathbf{w}|| <= -1$ if $y = 0$

Where **p** is projection of **x** onto **w**

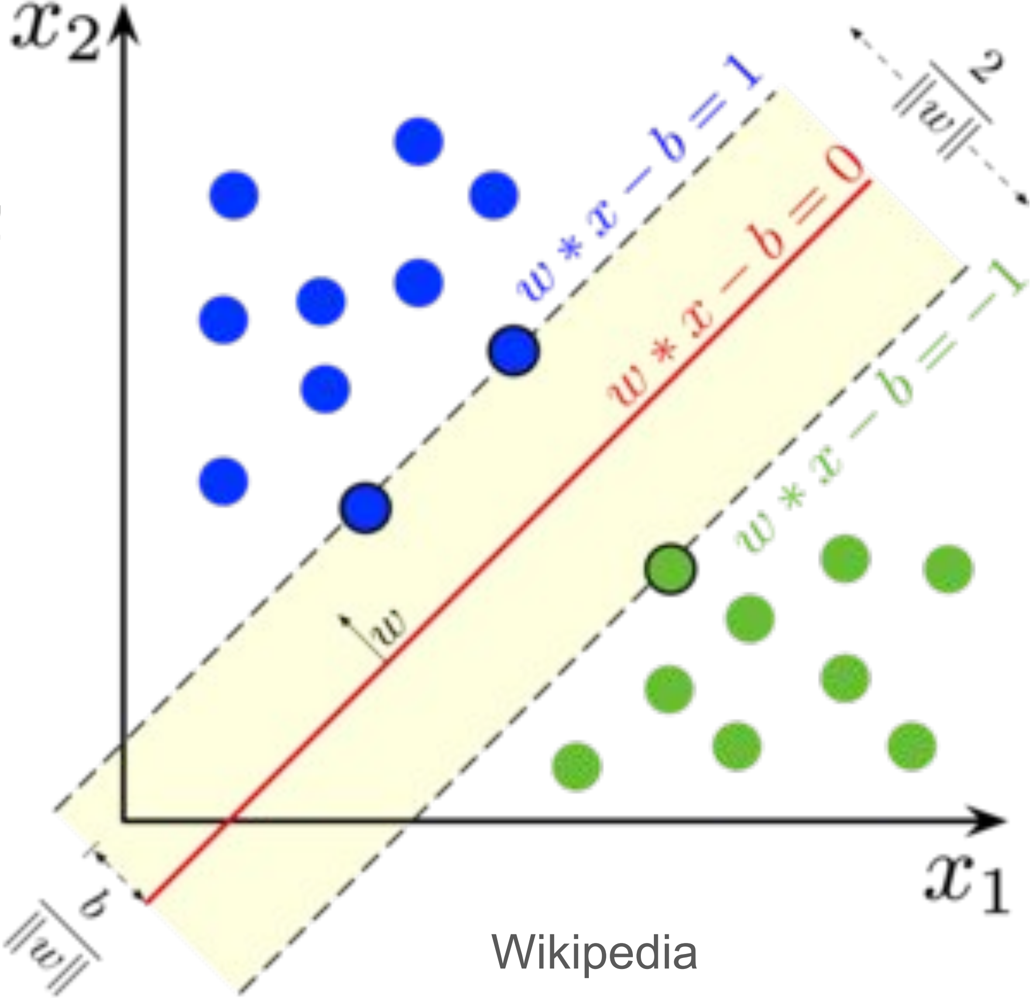
SVM Classification

Wikipedia



SVM Classification

Kernels will not be on the midterm!



Wikipedia