Support Vector Machines

Stephan Robert

1

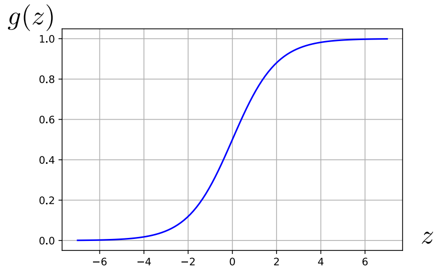
Last chapter: Logistic regression

- Linear regression

$$f_{\mathbf{\Theta}}(\mathbf{x}) = f(\mathbf{x}) = \mathbf{\Theta}^t\mathbf{x}$$

- Logistic regression

$$h_{\mathbf{\Theta}}(\mathbf{x}) = h(\mathbf{x}) = g(\mathbf{\Theta}^t\mathbf{x})$$



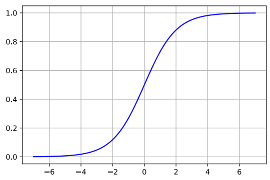
$$= \frac{1}{1+e^{-\mathbf{\Theta}^t\mathbf{x}}}$$

$\mathbf{\Theta}^t\mathbf{x} \in \mathbb{R}$

$z = \mathbf{\Theta}^t\mathbf{x}$

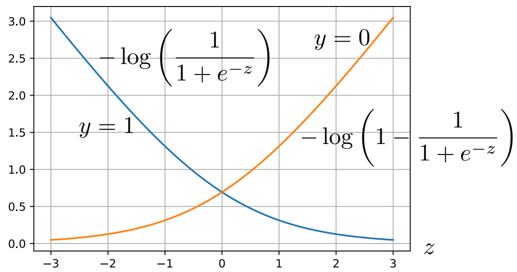
Sigmoid or Logistic function

Alternative view



$z = \mathbf{\Theta}^t \mathbf{x}$

Alternative view of logistic regression (cost as a function of z instead of $h_{\Theta}(\mathbf{x})$ ($h_{\Theta}(\mathbf{x}) \neq z$!)) :

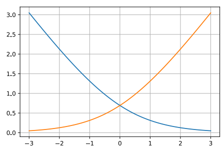


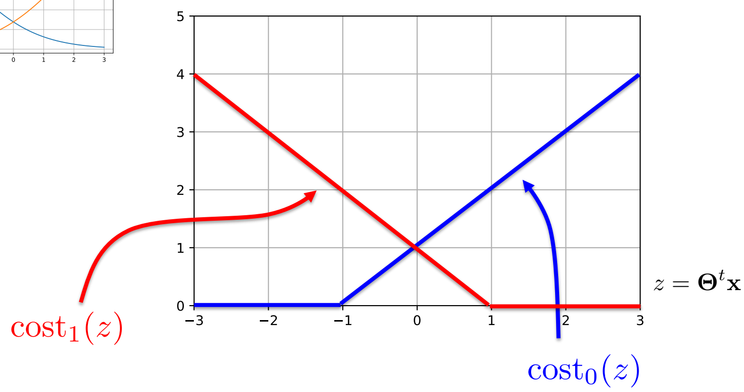
if $y = 1$ and $z = \mathbf{\Theta}^t\mathbf{x} \gg 0$ we want $h(\mathbf{x}) \approx 1$
if $y = 0$ and $z = \mathbf{\Theta}^t\mathbf{x} \ll 0$ we want $h(\mathbf{x}) \approx 0$

$$-y\log\left(\frac{1}{1+e^{-\mathbf{\Theta}^t\mathbf{x}}}\right) - (1-y)\log\left(1 - \frac{1}{1+e^{-\mathbf{\Theta}^t\mathbf{x}}}\right)$$

3

Support Vector machine





4

Cost functions

- Logistic Regression

$$J(\boldsymbol{\Theta}) = -\frac{1}{m}\left(\sum_{i=1}^{m}(y)_i \log h((\mathbf{x})_i) + (1-(y)_i)\log(1-h((\mathbf{x})_i)\right) + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

- Support Vector machine

$$J(\boldsymbol{\Theta}) = \frac{1}{\cancel{m}}\left(\sum_{i=1}^{m}(y)_i\text{cost}_1(\boldsymbol{\Theta}^t\mathbf{x}_i) + (1-(y)_i)\text{cost}_0(\boldsymbol{\Theta}^t\mathbf{x}_i)\right) + \frac{\lambda}{2\cancel{m}}\sum_{j=1}^{n}\theta_j^2$$

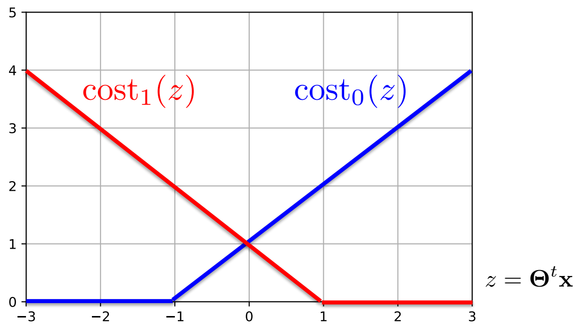
Important notice: By convention (SVM community) we get rid of m

5

Support Vector Machine

We want:

$$\min_{\boldsymbol{\Theta}\in\mathbb{R}^n} C\overbrace{\left(\sum_{i=1}^{m}(y)_i\text{cost}_1(\boldsymbol{\Theta}^t\mathbf{x}_i) + (1-(y)_i)\text{cost}_0(\boldsymbol{\Theta}^t\mathbf{x}_i)\right)}^{0} + \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$

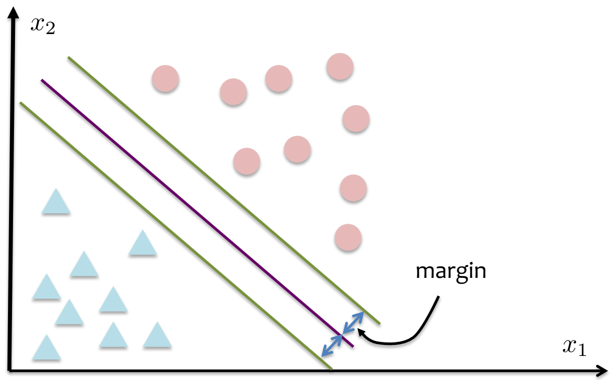


$$C = \frac{1}{\lambda}$$

6

SVM Decision Boundary

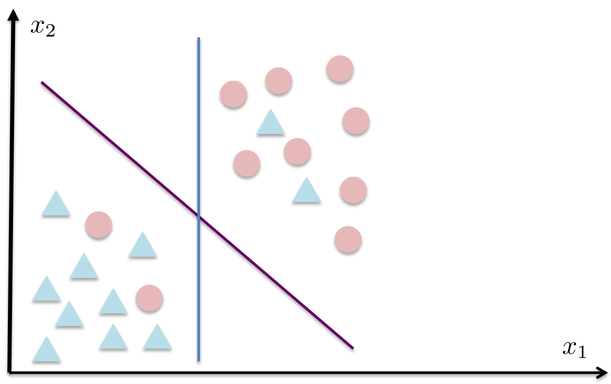
Linear Separable Case, Large margin Classifier



7

SVM Decision Boundary

Large margin in presence of outliers



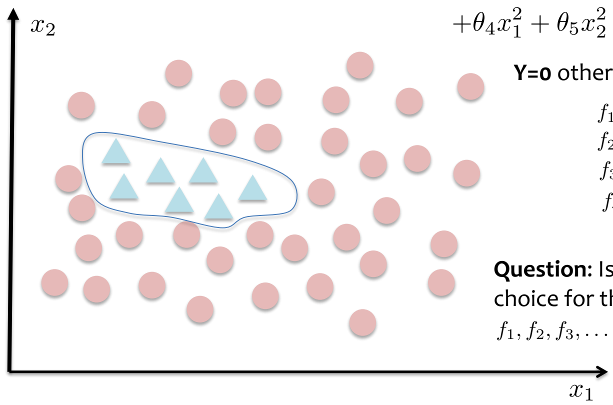
8

SVM Kernels

Non linear Decision Boundary

Predict **y=1** if $\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \ldots \geq 0$

Y=0 otherwise!!!



$f_1 = x_1$

$f_2 = x_2$

$f_3 = x_1 x_2$

$f_4 = x_1^2$

$\vdots$

Question: Is there a better choice for the features $f_1, f_2, f_3, \ldots$?

Predict $y = 1, h(\mathbf{x}) = 1$ if

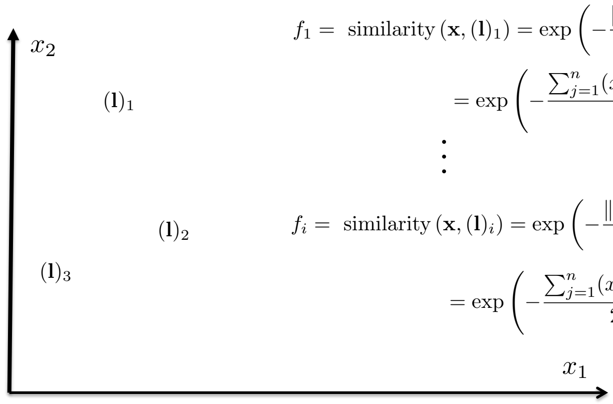
$$\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \ldots \geq 0$$

9

SVM Kernels

Non linear Decision Boundary

$\mathbf{x}$ is given



$$f_1 = \text{similarity}(\mathbf{x}, (\mathbf{l})_1) = \exp\left(-\frac{\|\mathbf{x} - (\mathbf{l})_1\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{j=1}^{n}(x_j - (l_j)_1)^2}{2\sigma^2}\right)$$

$$\vdots$$

$$f_i = \text{similarity}(\mathbf{x}, (\mathbf{l})_i) = \exp\left(-\frac{\|\mathbf{x} - (\mathbf{l})_i\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{j=1}^{n}(x_j - (l_j)_i)^2}{2\sigma^2}\right)$$

$$\mathbf{x} \approx (\mathbf{l})_1 \rightarrow f_1 \approx \exp\left(-\frac{0}{2\sigma^2}\right) \approx 1$$

$$\mathbf{x} \text{ far from } (\mathbf{l})_1 \rightarrow f_1 = \exp\left(-\frac{(\text{large number})^2}{2\sigma^2}\right) \approx 0$$

10

SVM Kernels

Non linear Decision Boundary

$\mathbf{x}$ is given

x_2

$(\mathbf{l})_1$

$\mathbf{x}$ and $(\mathbf{l})_1$ are near

$$f_1 = \text{similarity}\,(\mathbf{x}, (\mathbf{l})_1) = \exp\left(-\frac{\sum_{j=1}^{n}(x_j - (l_j)_1)^2}{2\sigma^2}\right) \approx 1$$

$\mathbf{x}$ and $(\mathbf{l})_1$ are far

$$f_1 = \text{similarity}\,(\mathbf{x}, (\mathbf{l})_1) = \exp\left(-\frac{\sum_{j=1}^{n}(x_j - (l_j)_1)^2}{2\sigma^2}\right) \approx 0$$

$(\mathbf{l})_2$

$(\mathbf{l})_3$

$$f_1 = \text{similarity}\,(\mathbf{x}, (\mathbf{l})_1) = \text{kernel}\,(\mathbf{x}, (\mathbf{l})_1)$$

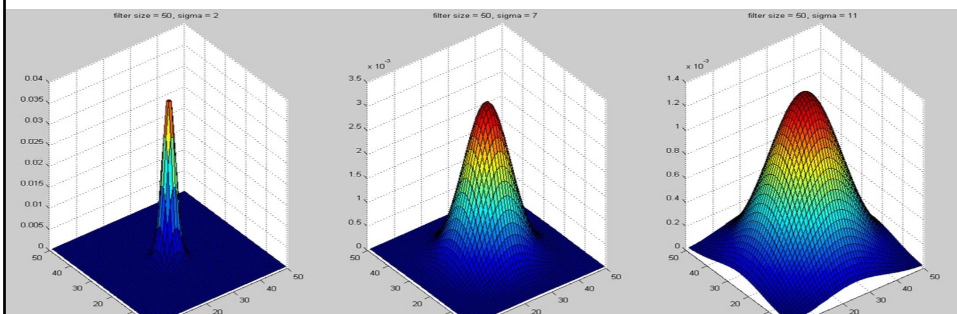
x_1

$$\mathbf{x} \approx (\mathbf{l})_1 \rightarrow f_1 \approx \exp\left(-\frac{0}{2\sigma^2}\right) \approx 1$$

$$\mathbf{x} \text{ far from } (\mathbf{l})_1 \rightarrow f_1 = \exp\left(-\frac{(\text{large number})^2}{2\sigma^2}\right) \approx 0$$

11

Gauss



12

SVM Kernels

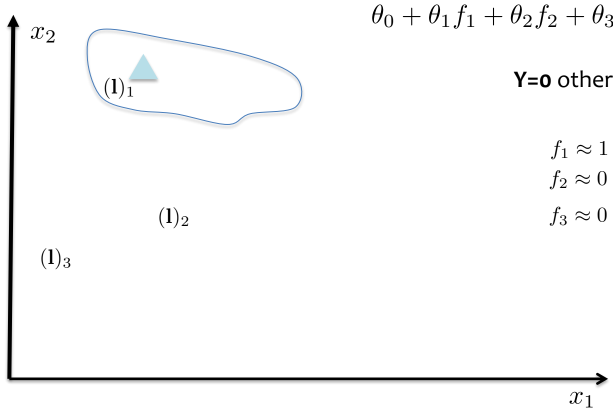
Non linear Decision Boundary

Predict **y=1** if

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$

Y=0 otherwise!!!

$$f_1 \approx 1$$
$$f_2 \approx 0$$
$$f_3 \approx 0$$



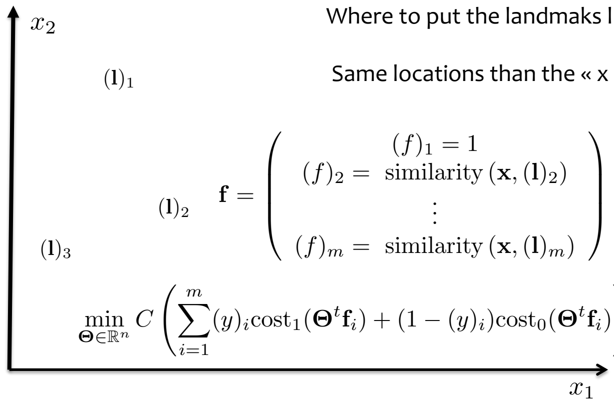
13

SVM Kernels

Non linear Decision Boundary

Where to put the landmaks l ?

Same locations than the « x » (training)



$$\mathbf{f} = \begin{pmatrix} (f)_1 = 1 \\ (f)_2 = \text{ similarity } (\mathbf{x}, (\mathbf{l})_2) \\ \vdots \\ (f)_m = \text{ similarity } (\mathbf{x}, (\mathbf{l})_m) \end{pmatrix} = \begin{pmatrix} (f)_1 \\ (f)_2 \\ \vdots \\ (f)_m \end{pmatrix}$$

$$\min_{\mathbf{\Theta} \in \mathbb{R}^n} C \left(\sum_{i=1}^{m} (y)_i \text{cost}_1(\mathbf{\Theta}^t \mathbf{f}_i) + (1 - (y)_i) \text{cost}_0(\mathbf{\Theta}^t \mathbf{f}_i) \right) + \frac{1}{2} \sum_{j=1}^{n=m} \theta_j^2$$

14