Machine Learning and Pervasive Computing

Stephan Sigg

Georg-August-University Goettingen, Computer Networks

07.01.2015

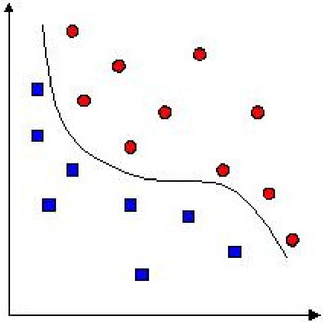
Overview and Structure

22.10.2014	Organisation
22.10.3014	Introduction (Def.: Machine learning, Supervised/Unsupervised, Examples)
29.10.2014	Machine Learning Basics (Toolchain, Features, Metrics, Rule-based)
05.11.2014	A simple Supervised learning algorithm
12.11.2014	Excursion: Avoiding local optima with random search
19.11.2014	–
26.11.2014	Bayesian learner
03.12.2014	–
10.12.2014	Decision tree learner
17.12.2014	k-nearest neighbour
07.01.2015	Support Vector Machines
14.01.2015	Artificial Neural networks and Self Organizing Maps
21.01.2015	Hidden Markov models and Conditional random fields
28.01.2015	High dimensional data, Unsupervised learning
04.02.2015	Anomaly detection, Online learning, Recom. systems

Outline

Support vector machines (SVM)

For our previous classifier, we have designed an objective function of sufficient dimension

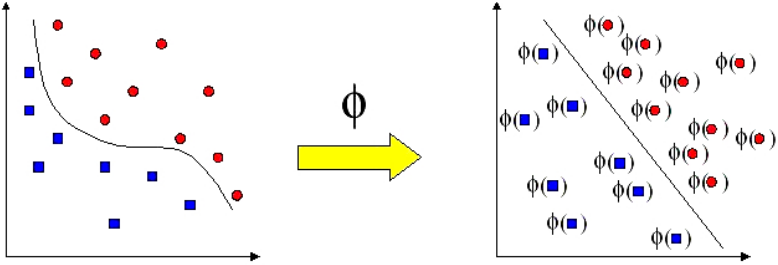


Support vector machines (SVM)

For our previous classifier, we have designed an objective function of sufficient dimension

Alternative to designing complex non-linear functions:

Change dimension of input space so that linear separator is again possible



Support vector machines (SVM)

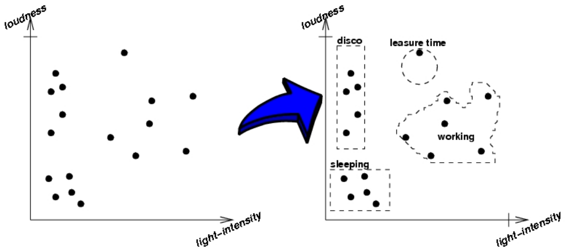
In classification we assign $\vec{x}$ to one of K discrete classes C_k

The input is divided by decision boundaries

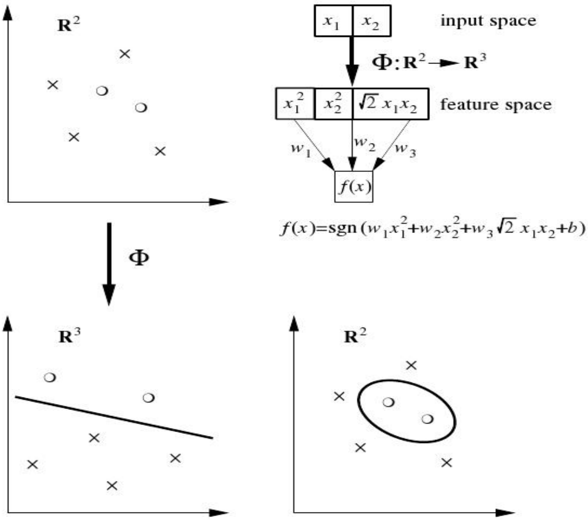
Here we assume that decision boundaries are linear functions of $\vec{x}$

Data separable by linear decision surfaces are linear separable

With high dimension, a set of two classes is always linear separable



Support vector machines (SVM)

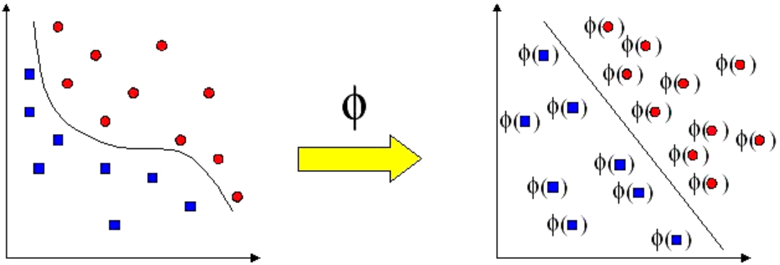


Support vector machines (SVM)

SVM pre-processes data to represent patterns in a high dimension

Dimension often much higher than original feature space

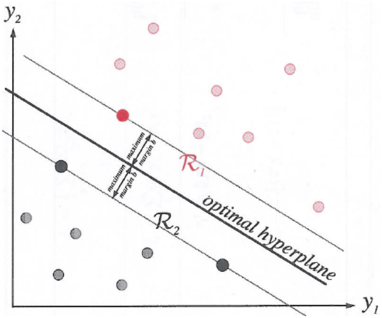
Then, insert hyperplane in order to separate the data



Support vector machines (SVM)

The goal for support vector machines is to find a separating hyperplane with the largest margin to the outer points in all sets

If no such hyperplane exists, map all points into a higher dimensional space until such a plane exists



Support vector machines (SVM)

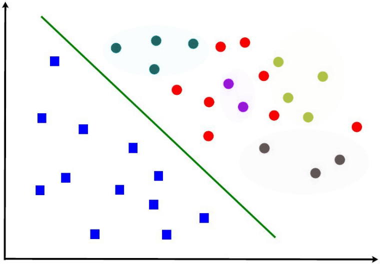
Simple application to several classes by iterative approach:

belongs to class 1 or not?

belongs to class 2 or not?

...

Search for optimum mapping between input space and feature space complicated (no optimum approach known)



Support vector machines (SVM)

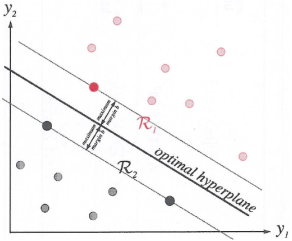
Simple learning approach to find the correct hyperplane:

Starting from an initial separating hyperplane

Find worst classified pattern (on the wrong side of the hyperplane)

Design a new hyperplane with this pattern as one of the support vectors

Iterate until all patterns are correctly classified



Outline

Support vector machines (SVM)

Cost function

Contribution of a single sample to the overall cost:

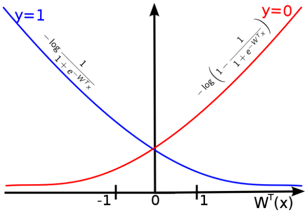
Support vector machines (SVM)

Cost function

Contribution of a single sample to the overall cost:

Logistic regression

$$-y \cdot \log \frac{1}{1+e^{-W^T x}} - (1-y) \cdot \log \left(1 - \frac{1}{1+e^{-W^T x}}\right)$$



Support vector machines (SVM)

Cost function

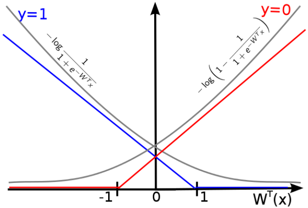
Contribution of a single sample to the overall cost:

Logistic regression

$$-y \cdot \log \frac{1}{1+e^{-W^T x}} - (1-y) \cdot \log \left(1 - \frac{1}{1+e^{-W^T x}}\right)$$

SVM

$$-y \cdot \text{cost}_{y=1}(W^T x) + -(1-y) \cdot \text{cost}_{y=0}(W^T x)$$



Support vector machines (SVM)

Cost function

Logistic regression

$$\min_{W} \frac{1}{m}\left[\sum_{i=1}^{m} y_i\left(-\log\frac{1}{1+e^{-W^T x_i}}\right)+(1-y_i)\left(-\log\left(1-\frac{1}{1+e^{-W^T x_i}}\right)\right)\right]+\frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

SVM

$$\min_{W} C\sum_{i=1}^{m}\left[y_i \mathrm{cost}_{y=1}(W^T x_i)+(1-y_i)\mathrm{cost}_{y=0}(W^T x_i)\right]+\frac{1}{2}\sum_{j=1}^{n} w_j^2$$

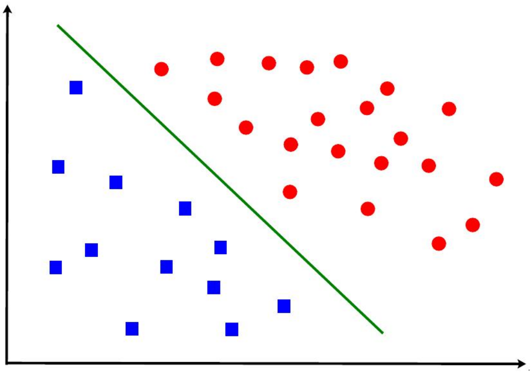
1

[1] C here plays a similar role as $\frac{1}{\lambda}$

Outline

Support vector machines (SVM)

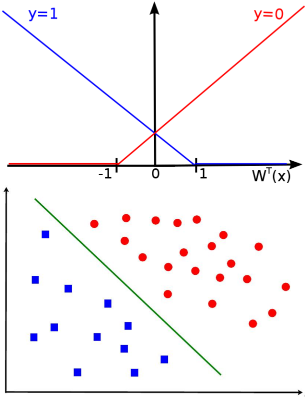
SVM hypothesis



$$\min_{W} C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

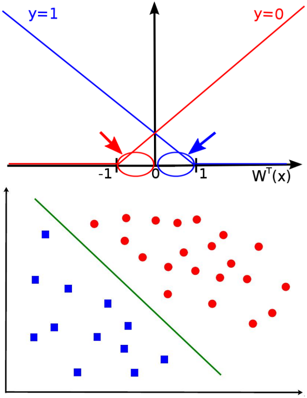
SVM hypothesis



$$\min_{W} C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

SVM hypothesis

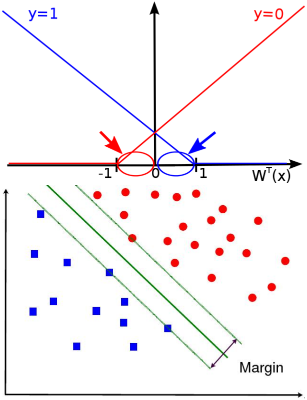


$$W^T x \begin{cases} \geq 0 \\ < 0 \end{cases} \text{ sufficient}$$

$$\min_W C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

SVM hypothesis

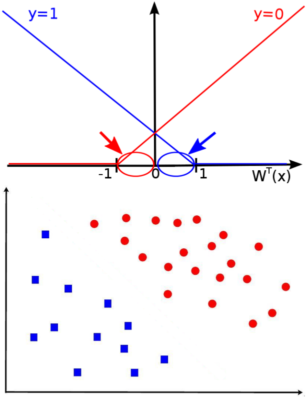


$$W^T x \begin{cases} \geq 1 \\ \leq -1 \end{cases} \Rightarrow \text{confidence}$$

$$\min_W C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

SVM hypothesis



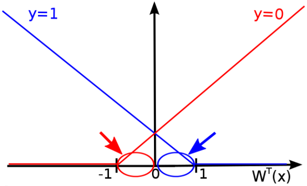
$$W^T x \begin{cases} \geq 1 \\ \leq -1 \end{cases} \Rightarrow \text{confidence}$$

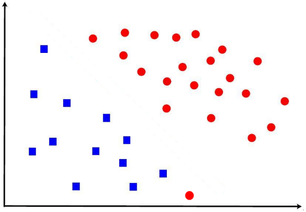
Outliers: Elastic decision boundary

$$\min_W C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

SVM hypothesis





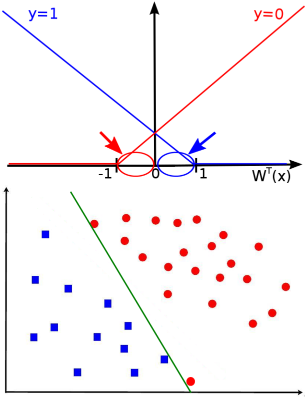
$$W^T x \begin{cases} \geq 1 \\ \leq -1 \end{cases} \Rightarrow \text{confidence}$$

Outliers: Elastic decision boundary

$$\min_W C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

SVM hypothesis



$$W^T x \begin{cases} \geq \; 1 \\ \leq -1 \end{cases} \Rightarrow \text{confidence}$$

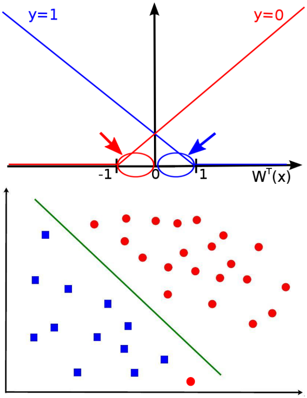
Outliers: Elastic decision boundary

large C stricter boundary at the cost of smaller margin

$$\min_W C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

SVM hypothesis



$$W^T x \begin{cases} \geq 1 \\ \leq -1 \end{cases} \Rightarrow \text{confidence}$$

Outliers: Elastic decision boundary

small C tolerates outliers

$$\min_W C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

Large margin classifier

$$\min_W C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Support vector machines (SVM)

Large margin classifier

$$\min_{W} C \sum_{i=1}^{m} \left[y_i \text{cost}_{y=1}(W^T x_i) + (1 - y_i)\text{cost}_{y=0}(W^T x_i) \right] + \frac{1}{2} \sum_{j=1}^{n} w_j^2$$

Rewrite the SVM optimisation problem as

$$\begin{aligned} \min_{W} \quad & \tfrac{1}{2} \textstyle\sum_{j=1}^{n} w_j^2 \\ s.t. \quad & W^T x_i \geq 1 \quad \text{if } y_i = 1 \\ & W^T x_i \leq -1 \quad \text{if } y_i = 0 \end{aligned}$$

Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2$$

$$\begin{aligned} s.t. \quad & W^T x_i \geq 1 \ \text{ if } y_i = 1 \\ & W^T x_i \leq -1 \text{ if } y_i = 0 \end{aligned}$$

Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2$$

$$s.t. \quad W^T x_i \geq 1 \ \text{ if } y_i = 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0$$

Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 = \frac{1}{2}||W||^2$$

$$s.t. \qquad W^T x_i \geq 1 \;\text{ if } y_i = 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0$$

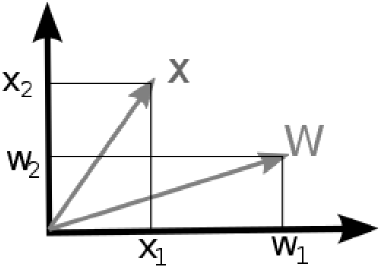
Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 = \frac{1}{2}||W||^2$$

$$s.t. \qquad W^T x_i \geq 1 \;\text{ if } y_i = 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0$$



$$W^T x = w_1 x_1 + w_2 x_2$$

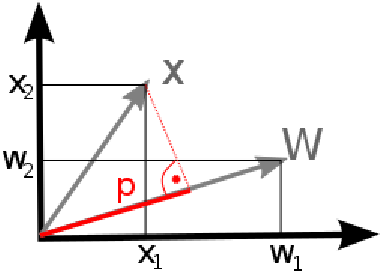
Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 = \frac{1}{2}||W||^2$$

$$s.t. \qquad W^T x_i \geq 1 \;\text{ if } y_i = 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0$$



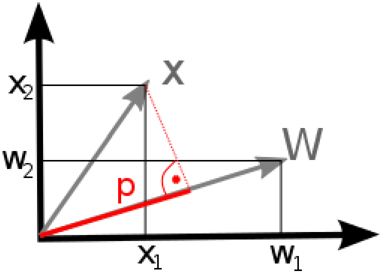
$$W^T x = w_1 x_1 + w_2 x_2 = ||W|| \cdot p$$

Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2+\cdots+w_n^2}\right)^2 \quad = \frac{1}{2}||W||^2$$

$$\begin{aligned} s.t. \quad & W^T x_i \geq 1 \text{ if } y_i = 1 && \rightarrow ||W|| \cdot p_i \geq 1 \\ & W^T x_i \leq -1 \text{ if } y_i = 0 && \rightarrow ||W|| \cdot p_i \leq -1 \end{aligned}$$



$$W^T x = w_1 x_1 + w_2 x_2 = ||W|| \cdot p$$

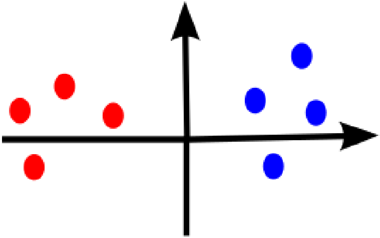
Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 \quad = \frac{1}{2}||W||^2$$

$$s.t. \quad W^T x_i \geq 1 \;\text{ if } y_i = 1 \quad \rightarrow ||W|| \cdot p_i \geq 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0 \quad \rightarrow ||W|| \cdot p_i \leq -1$$



Which decision boundaray is found?

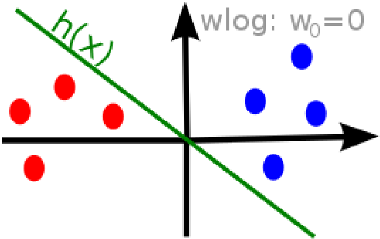
Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 \quad = \frac{1}{2}||W||^2$$

$$s.t. \qquad W^T x_i \geq 1 \text{ if } y_i = 1 \qquad \rightarrow ||W|| \cdot p_i \geq 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0 \qquad \rightarrow ||W|| \cdot p_i \leq -1$$



Which decision boundaray is found?

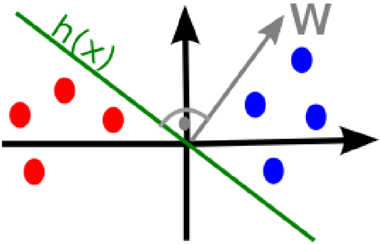
Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 \quad = \frac{1}{2}||W||^2$$

$$s.t. \quad W^T x_i \geq 1 \text{ if } y_i = 1 \quad \rightarrow ||W|| \cdot p_i \geq 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0 \quad \rightarrow ||W|| \cdot p_i \leq -1$$



Which decision boundaray is found?

$$h(x) = w_1 x_1 + w_2 x_2$$

$\rightarrow$ W orthogonal to all x with $h(x) = 0$

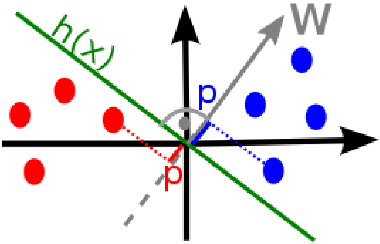
Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 \quad = \frac{1}{2}||W||^2$$

$$s.t. \quad W^T x_i \geq 1 \text{ if } y_i = 1 \quad \rightarrow ||W|| \cdot p_i \geq 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0 \quad \rightarrow ||W|| \cdot p_i \leq -1$$



Which decision boundaray is found?

$$h(x) = w_1 x_1 + w_2 x_2$$

$\rightarrow$ W orthogonal to all x with $h(x) = 0$

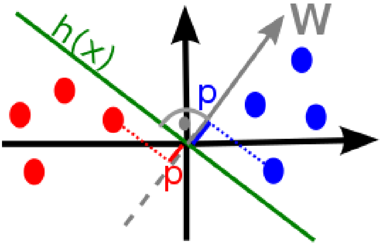
Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 = \frac{1}{2}||W||^2$$

$$s.t. \quad W^T x_i \geq 1 \text{ if } y_i = 1 \quad \rightarrow ||W|| \cdot p_i \geq 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0 \quad \rightarrow ||W|| \cdot p_i \leq -1$$



Which decision boundaray is found?

$h(x) = w_1 x_1 + w_2 x_2$

$\rightarrow$ W orthogonal to all x with $h(x) = 0$

$\Rightarrow$ min $\frac{1}{2}||W||^2$ and $||W|| \cdot p_i \geq 1$ necessitate larger p_i

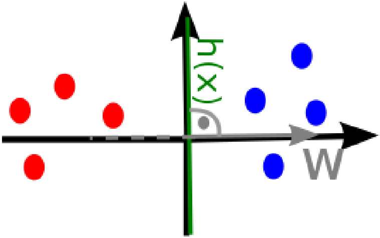
Support vector machines (SVM)

Large margin classifier

$$\min_{W} \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 = \frac{1}{2}||W||^2$$

$$s.t. \quad W^T x_i \geq 1 \;\text{ if } y_i = 1 \quad \rightarrow ||W|| \cdot p_i \geq 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0 \quad \rightarrow ||W|| \cdot p_i \leq -1$$



Which decision boundaray is found?

$h(x) = w_1 x_1 + w_2 x_2$

$\rightarrow$ W orthogonal to all x with $h(x) = 0$

$\Rightarrow$ $\min \frac{1}{2}||W||^2$ and $||W|| \cdot p_i \geq 1$ necessitate larger p_i

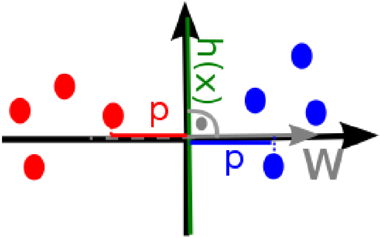
Support vector machines (SVM)

Large margin classifier

$$\min_W \quad \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \frac{1}{2}\left(\sqrt{w_1^2 + \cdots + w_n^2}\right)^2 \quad = \frac{1}{2}||W||^2$$

$$s.t. \quad W^T x_i \geq 1 \;\text{ if } y_i = 1 \quad \rightarrow ||W|| \cdot p_i \geq 1$$

$$W^T x_i \leq -1 \text{ if } y_i = 0 \quad \rightarrow ||W|| \cdot p_i \leq -1$$



Which decision boundaray is found?

$h(x) = w_1 x_1 + w_2 x_2$

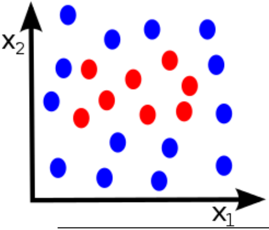
$\rightarrow$ W orthogonal to all x with $h(x) = 0$

$\Rightarrow$ $\min \frac{1}{2}||W||^2$ and $||W|| \cdot p_i \geq 1$ necessitate larger p_i

Outline

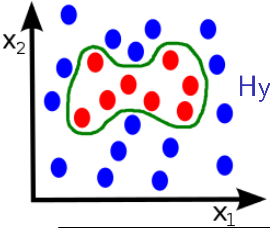
Support vector machines (SVM)

Kernels – Non linear decision boundary



Support vector machines (SVM)

Kernels – Non linear decision boundary

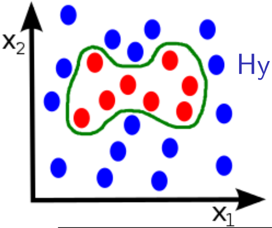


Hypothesis $= 1$ if

$$w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 + w_4 x_1^2 + \cdots \geq 0$$

Support vector machines (SVM)

Kernels – Non linear decision boundary



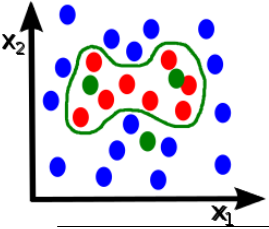
Hypothesis $= 1$ if

$$w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 + w_4 x_1^2 + \cdots \geq 0$$

$$\Rightarrow w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \ldots$$

Support vector machines (SVM)

Kernels – Non linear decision boundary

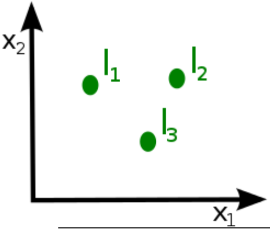


$$\Rightarrow w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \dots$$

Kernel Define kernel via landmarks

Support vector machines (SVM)

Kernels – Non linear decision boundary

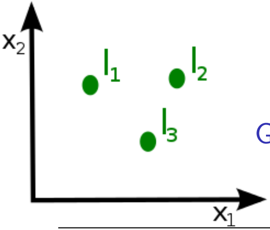


$$\Rightarrow \; w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \ldots$$

Kernel Define kernel via landmarks

Support vector machines (SVM)

Kernels – Non linear decision boundary

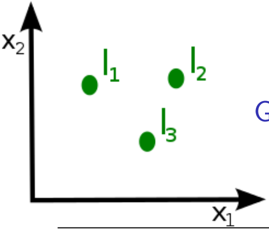


$$\Rightarrow w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \dots$$

Gaussian: $k(x, l_i) = e^{-\frac{||x-l_i||^2}{2\sigma^2}}$

Support vector machines (SVM)

Kernels – Non linear decision boundary



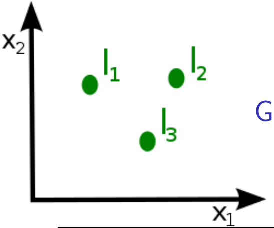
$$\Rightarrow w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \ldots$$

Gaussian: $k(x, l_i) = e^{-\frac{||x-l_i||^2}{2\sigma^2}}$

$x \approx l_i \Rightarrow k(x, l_i) \approx 1$ (0 else)

Support vector machines (SVM)

Kernels – Non linear decision boundary



$$\Rightarrow w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \ldots$$

Gaussian: $k(x, l_i) = e^{-\frac{||x - l_i||^2}{2\sigma^2}}$

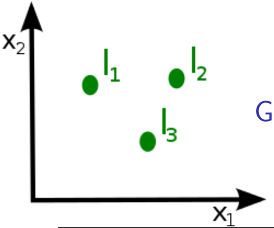
$x \approx l_i \Rightarrow k(x, l_i) \approx 1$ (0 else)

Example: $w_0 = -0.5, w_1 = 1, w_2 = 1, w_3 = 0$

$$h(x) = w_0 + w_1 k(x, l_1) + w_2 k(x, l_2) + w_3 k(x, l_3)$$

Support vector machines (SVM)

Kernels – Non linear decision boundary



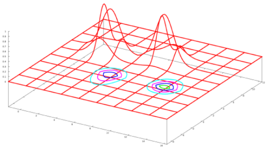
$$\Rightarrow w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \ldots$$

Gaussian: $k(x, l_i) = e^{-\frac{||x-l_i||^2}{2\sigma^2}}$

$x \approx l_i \Rightarrow k(x, l_i) \approx 1$ (0 else)

Example: $w_0 = -0.5, w_1 = 1, w_2 = 1, w_3 = 0$

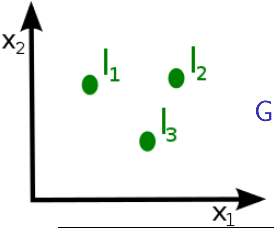
$$h(x) = w_0 + w_1 k(x, l_1) + w_2 k(x, l_2) + w_3 k(x, l_3)$$



$\sigma = 1$

Support vector machines (SVM)

Kernels – Non linear decision boundary



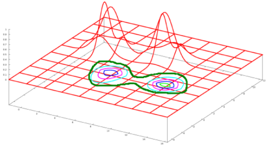
$$\Rightarrow \; w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \ldots$$

Gaussian: $k(x, l_i) = e^{-\frac{||x - l_i||^2}{2\sigma^2}}$

$x \approx l_i \;\Rightarrow\; k(x, l_i) \approx 1$ (0 else)

Example: $w_0 = -0.5, w_1 = 1, w_2 = 1, w_3 = 0$

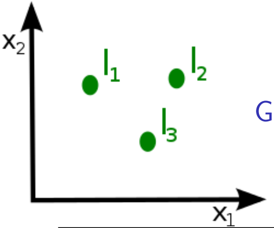
$$h(x) = w_0 + w_1 k(x, l_1) + w_2 k(x, l_2) + w_3 k(x, l_3)$$



$\sigma = 1$

Support vector machines (SVM)

Kernels – Non linear decision boundary



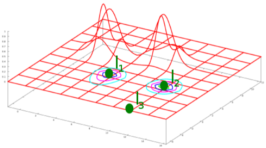
$$\Rightarrow \; w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \dots$$

Gaussian: $k(x, l_i) = e^{-\frac{||x - l_i||^2}{2\sigma^2}}$

$x \approx l_i \; \Rightarrow k(x, l_i) \approx 1$ (0 else)

Example: $w_0 = -0.5, w_1 = 1, w_2 = 1, w_3 = 0$

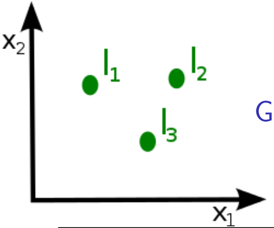
$$h(x) = w_0 + w_1 k(x, l_1) + w_2 k(x, l_2) + w_3 k(x, l_3)$$



$\sigma = 1$

Support vector machines (SVM)

Kernels – Non linear decision boundary



$$\Rightarrow \; w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \ldots$$

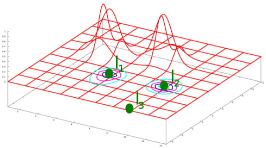
Gaussian: $k(x, l_i) = e^{-\frac{||x-l_i||^2}{2\sigma^2}}$

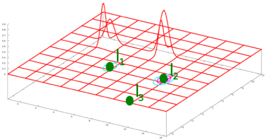
$x \approx l_i \Rightarrow k(x, l_i) \approx 1$ (0 else)

σ controls the width of the Gaussian

Example: $w_0 = -0.5, w_1 = 1, w_2 = 1, w_3 = 0$

$$h(x) = w_0 + w_1 k(x, l_1) + w_2 k(x, l_2) + w_3 k(x, l_3)$$

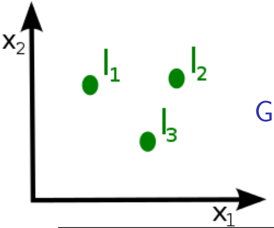




$\sigma = 1$ $\sigma = 0.5$

Support vector machines (SVM)

Kernels – Non linear decision boundary



$$\Rightarrow \; w_0 + w_1 k_1 + w_2 k_2 + w_3 k_3 + \ldots$$

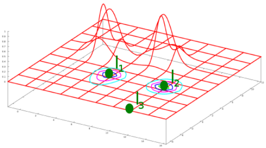
Gaussian: $k(x, l_i) = e^{-\frac{||x - l_i||^2}{2\sigma^2}}$

$x \approx l_i \;\Rightarrow k(x, l_i) \approx 1$ (0 else)

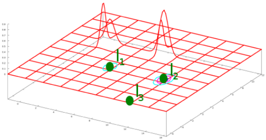
σ controls the width of the Gaussian

Example: $w_0 = -0.5, w_1 = 1, w_2 = 1, w_3 = 0$

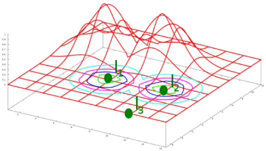
$$h(x) = w_0 + w_1 k(x, l_1) + w_2 k(x, l_2) + w_3 k(x, l_3)$$



$\sigma = 1$



$\sigma = 0.5$



$\sigma = 2$

Support vector machines (SVM)

Kernels – placement of landmarks

Possible choice of initial landmarks: All training-set samples

Training of w_i

$$f_i = \begin{bmatrix} k(x_i, l_1) \\ \vdots \\ k(x_i, l_m) \end{bmatrix}$$

$$\min_W C \sum_{i=1}^{m} y_i \text{cost}_{y_i=1}(W^T f_i) + (1 - y_i) \cdot \text{cost}_{y_i=0}(W^T f_i) + \frac{1}{2} \sum_{j=1}^{m} w_j^2$$

Outline

Questions?

Stephan Sigg
stephan.sigg@cs.uni-goettingen.de

Literature

- C.M. Bishop: Pattern recognition and machine learning, Springer, 2007.
- R.O. Duda, P.E. Hart, D.G. Stork: Pattern Classification, Wiley, 2001.