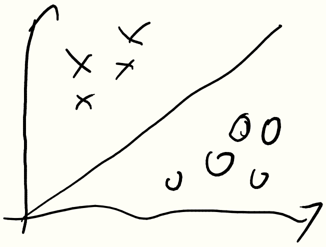
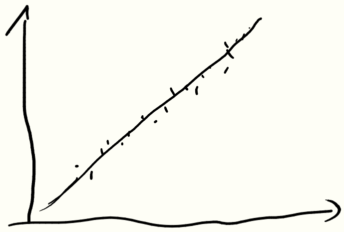
Classification



Regression



linear Model both methods

integer volues for y

contiunuum values for y

$$f(x,\theta) = \vec{x}\,\vec{\theta} + \theta_0$$ → Hyperplane or 2D line

loss-function Maß wie gut oder schlecht θ ist

$$L \qquad \underset{\theta}{\arg\min} \sum_{i}^{n} L(f(\theta,x_i),x_i) + \underbrace{\overset{\text{trade-off}}{\lambda}}\ \underbrace{\Omega(\theta)}_{\text{Regularisation term}}$$

stabilize numerical stabilization

reduce over fitting

finding Theta - use method gradient descent

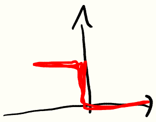
Algorithm
0. initialize theta
1. calculate gradient
2. calculate step size alpha ~ gradient
3. theta_n+1 = theta_n - alpha*gradient

problems:
more then one minima -> convex function
needs to much time -> stochastic gradient descent and parallel progr.

stochastic gradient:
use only a subset of the data for gradient calculation

loss-functions:

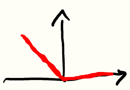
zero one loss



$$L = \begin{cases} 0, & y_i = \hat{y}_i \\ 1, & y_i \neq \hat{y}_i \end{cases}$$

not convex

perceptron loss



$$L = \begin{cases} 0, & y_i = \hat{y}_i \\ -y_i f(\theta,x_i), & y_i \neq \hat{y}_i \end{cases}$$

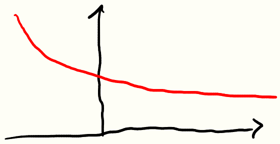
$$f(\theta,x) = \vec{x}\cdot\vec{\theta}$$

hinge loss

$$\max(0, 1 - y_i f(\theta,x_i))$$

logistic loss function

numeric stable



$$\log\left(1+e^{-y_i f(\theta, x_i)}\right)$$

$$\sum_i (\hat{y}_i - y_i)^2$$

or $\sum_i |\hat{y}_i - y_i|$

do not use for Regression modell the hinge loss

for classification use hinge loss

Regularizier - R - Ω

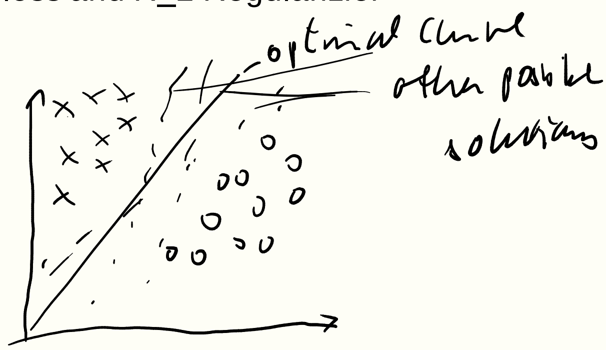
$\Omega_0(\theta) =$ Anzahl der nicht 0 Werte

$$\Omega_1(\theta) = \sum_i |\theta_i|$$

$$\Omega_2(\theta) = \sum_i \theta_i^2$$

Perceptron - linear classification method
with perceptron loss function

SVM - support vector machine
Hinge loss and R_2 Regularizier



algorithm tries to maximies the overall distance
to the cluster points, because there are many
possible solutions

$d = \frac{1}{\|\theta\|_2}$ distance

Model evaluation

independent identical distribution IID

empirical risc $R_S(\theta) = \frac{1}{n}\sum_i^n L(f(\theta, x_i))$ loss function

estimator expectation value $E(\theta) = \mu$

$B < 0$ pessimistic

$B > 0$ optimistic

bias $B = \mu - \theta_{real}$

variance $V = \sum_i |\mu - \theta_i|^2$

precision $P = \dfrac{n_{TP}}{n_{TP} + n_{FP}}$

recall $R = \dfrac{n_{TP}}{n_{TP} + n_{FN}}$

ROC curve