SciLifeLab

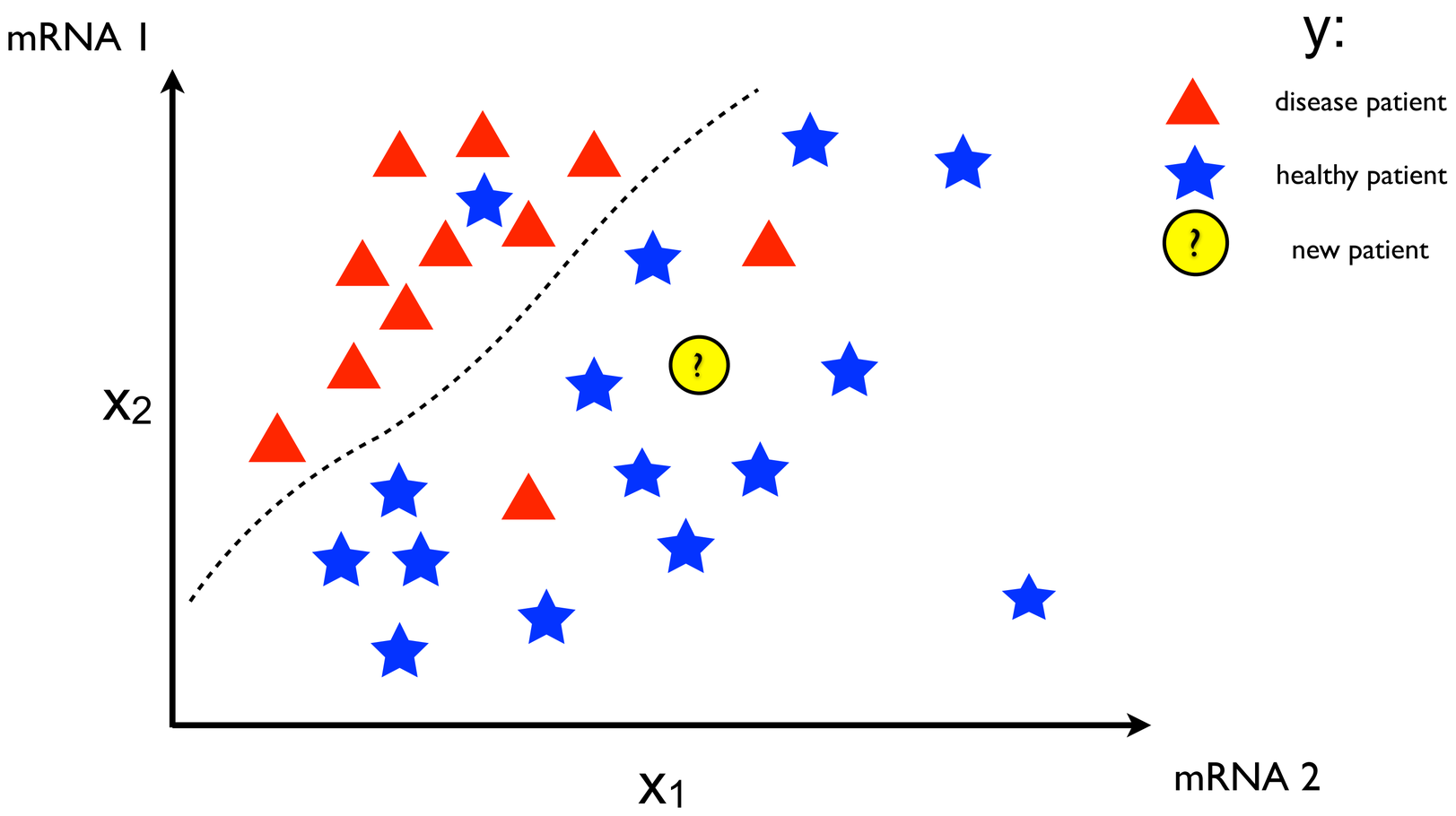
Supervised Machine Learning

CB2030

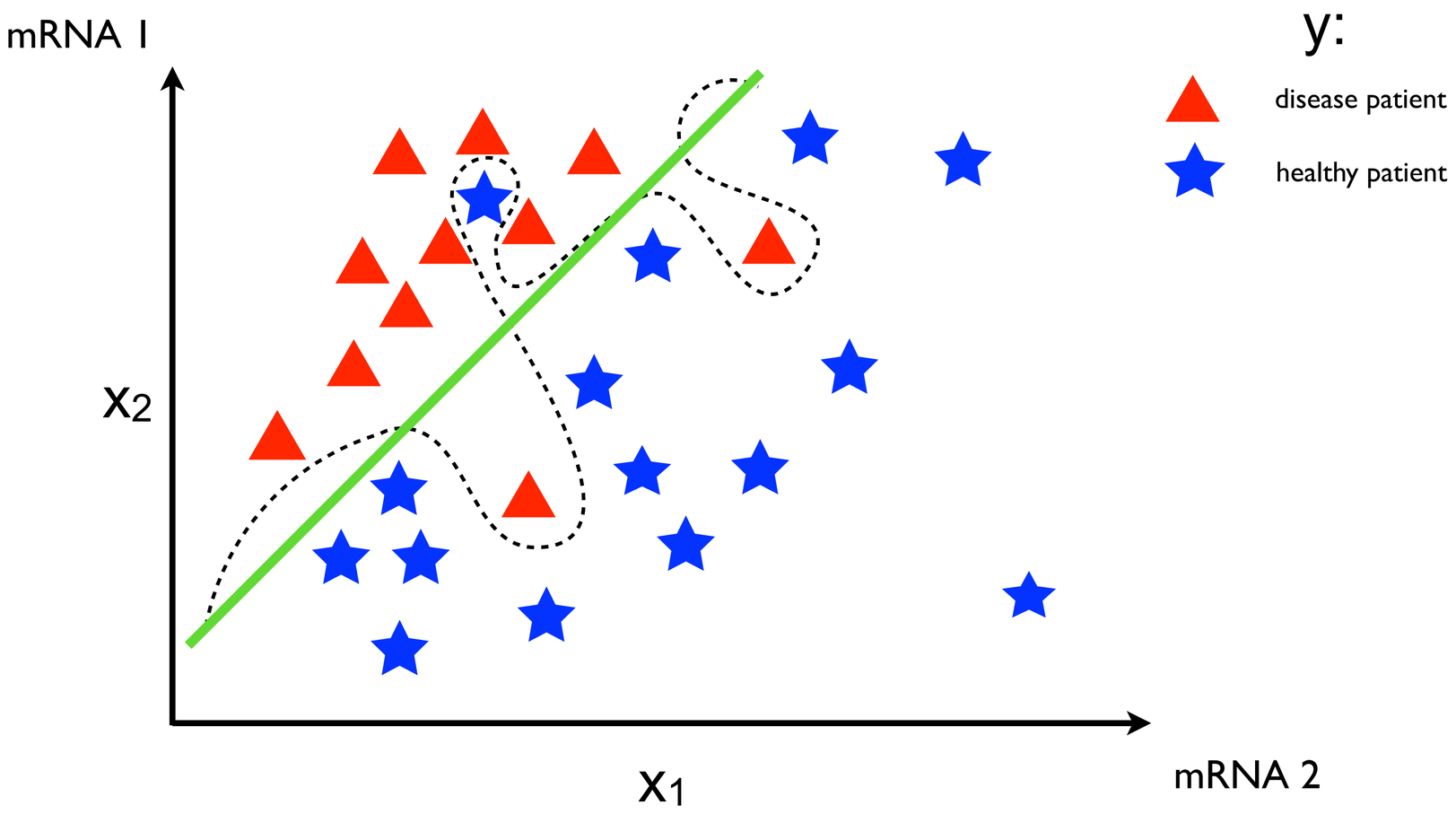
Lukas Käll, KTH

KTH VETENSKAP OCH KONST

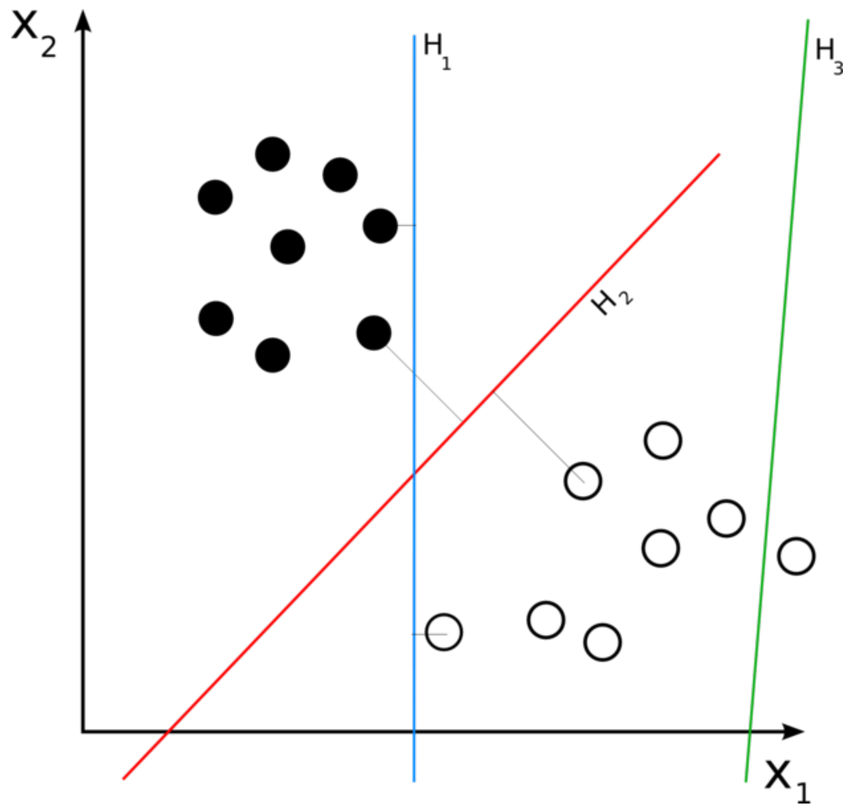
Supervised learning



Generalization: How to avoid



Separating hyperplane

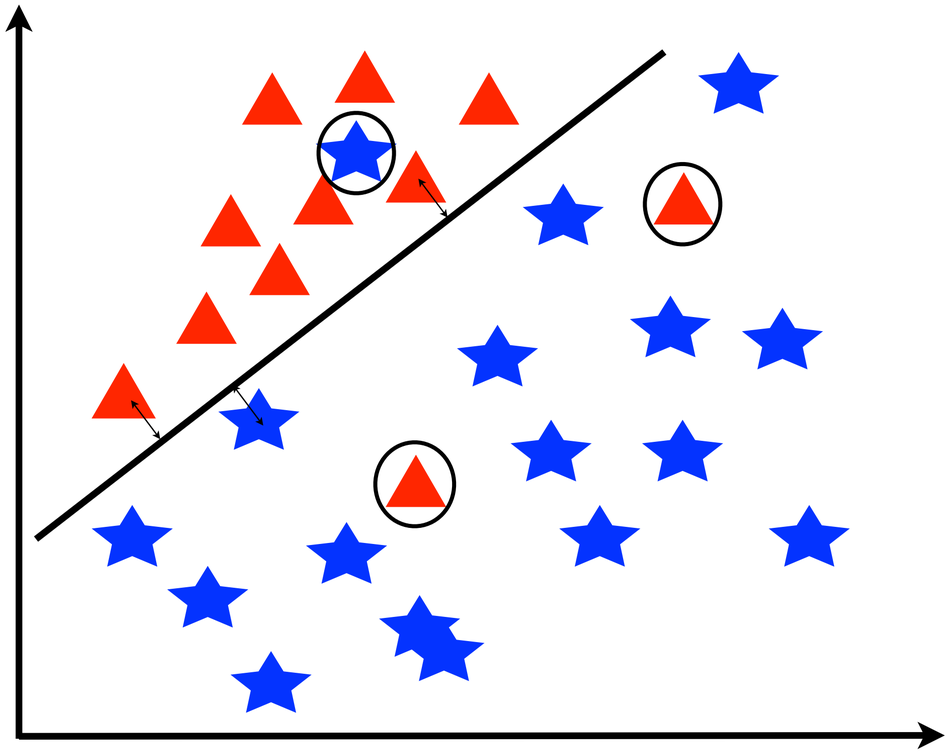


Example from Wikipedia

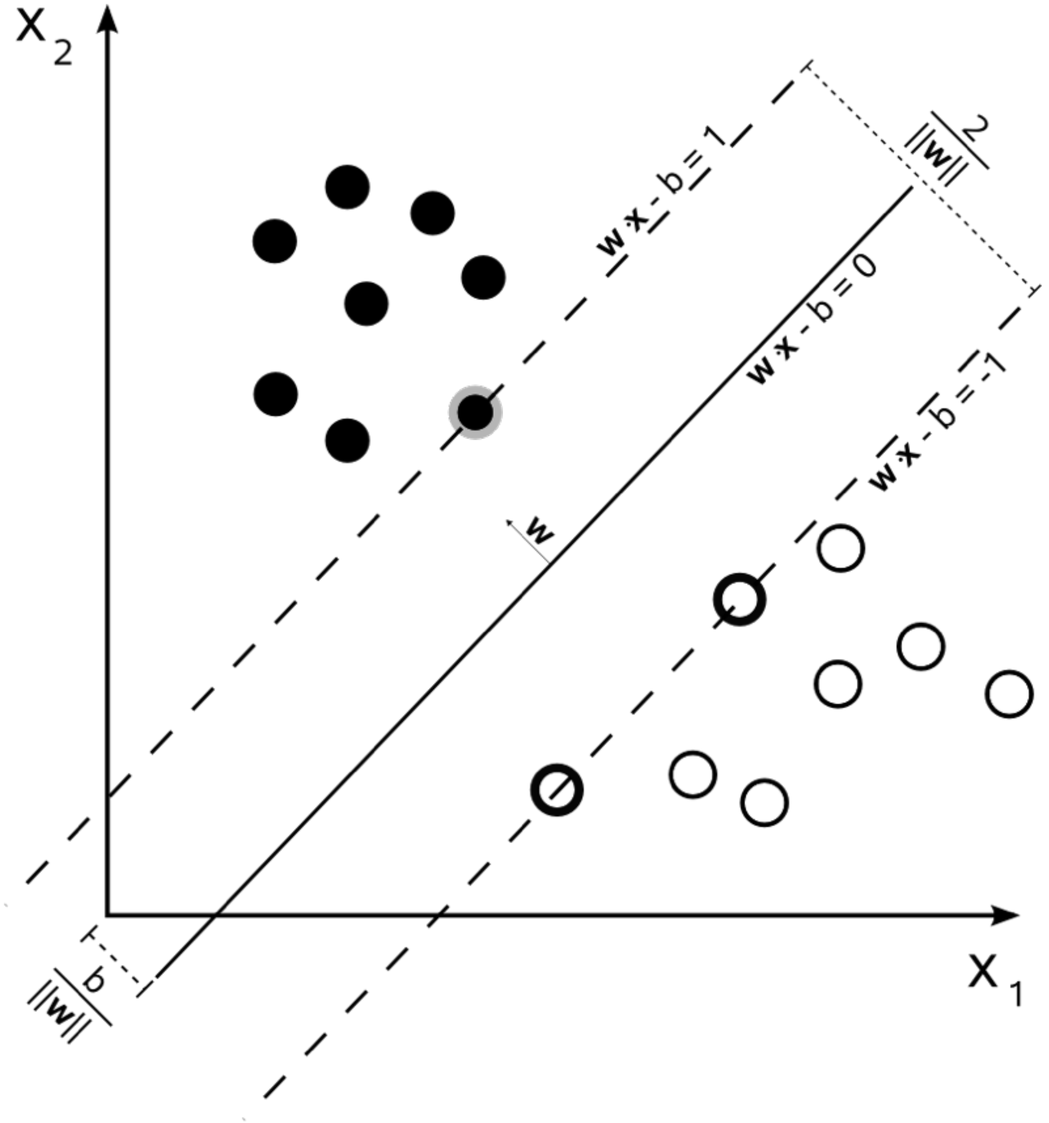
H3 (green) doesn't separate the two classes. H1 (blue) does, with a small margin and H2 (red) with the maximum margin.

Support Vector Machine

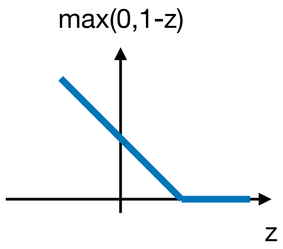
- Select a Maximum-margin separating hyper plane
- Soft margin, i.e. allow some data points to push their way through the margin of the separating hyperplane without affecting the end result too much
- Sometimes: transform your classification space using a kernel



Maximum margin hyperplane



Hinge loss function:



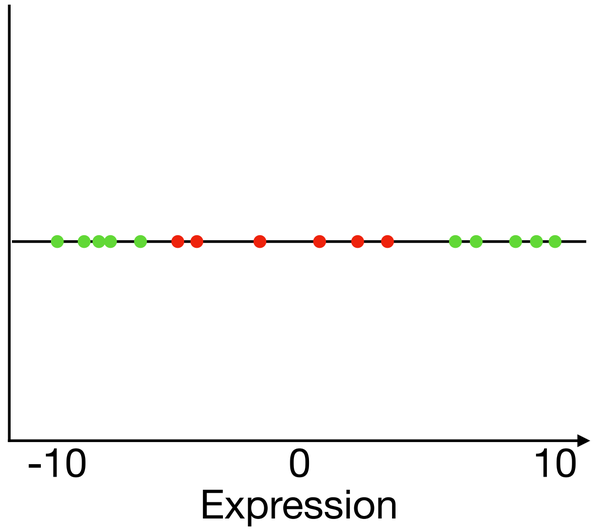
Minimize:

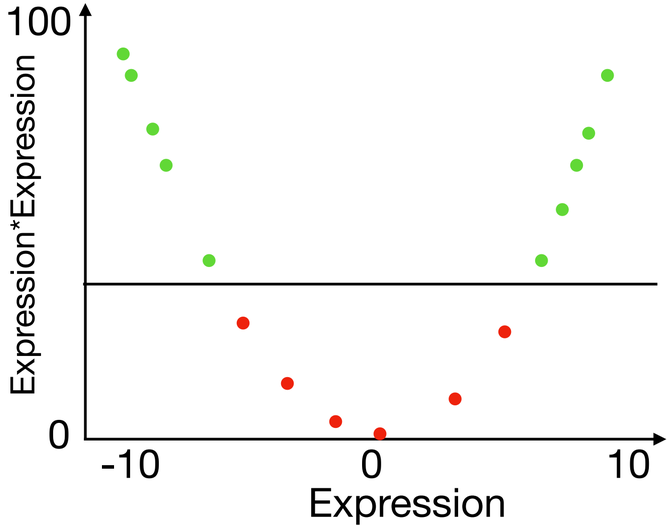
$$\left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0, 1 - y_i(\vec{w}\cdot\vec{x}_i - b)\right)\right] + \lambda\|\vec{w}\|^2,$$

Example from Wikipedia

Kernels

- Non-linear separation problem may be transformed into a linear problem if we select the right kernel





Example adopted from
[Noble 2006, Nat Biotech]

Strategies to validate supervised methods

- If we want to be able to detect over-fitting we need to train our method examples in a training set that is separate from the examples that we test our method with.
- If we need to optimise hyper-parameters we need to do so on yet another separate test.

Cross Validation

3-fold cross validation

Learner 1:

Learner 2:

Learner 3:

Nested Cross Validation

Learner 1:

Internal
X-val for L1:

Measuring performance of supervised classifiers

Score	Example type
7,5	+ Label
7,2	+ Label
5.0	+ Label
3,8	+ Label
3,7	- Label
2,5	+ Label
2,4	+ Label
1,4	+ Label
0,3	- Label
0,1	+ Label
-0,3	- Label
-1,4	+ Label
-2,3	- Label
-3,5	- Label
-4,4	+ Label
-5,3	- Label
-6,2	- Label

Predicted Positive

threshold

Predicted Negative

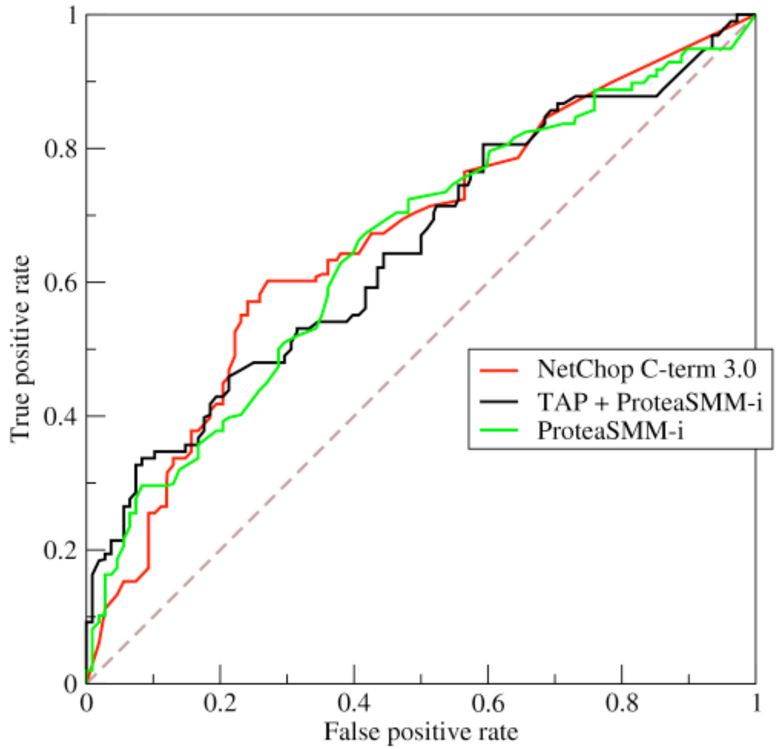
Performance metrics of supervised classifiers

	Predicted as positive	Predicted as negative
Positive example	TP	FN
Negative example	FP	TN

- TP = True positive =
  Correctly predicted as positive example
- FP = False positive =
  Incorrectly predicted as positive example
- FN = False negative =
  Incorrectly predicted as negative example
- TN = True negative =
  Correctly predicted as negative example

- Precision = TP/(TP+FP)
- Recall = Sensitivity = TP/(TP+FN)
- Specificity = TN/(TN+FP)
- FPR = FP/(FP+TN)
- TPR = TP/(TP+FN)
- FDR = FP/(FP+TP)

Receiver operating characteristic (ROC) plot



Example from
Wikipedia

ROC score =
area under the ROC curve