Testing and Error Estimation

Machine Learning

Prof Hans Georg Schaathun

Høgskolen i Ålesund

February 2, 2015

hials

Motivation

1. The perceptron
   - our first learning classifier
   - linear classifier
   - implmentation; reads data from file
2. Data set
   1. breast cancer data from Wisconsin
   2. two classes: benign and malign
3. Does the classifier work?
   - How do we test it?

hials

Outline

1 Error Estimation

2 Statistical Estimation

3 Evaluation Heuristic

4 Summary

hials

What do we measure?

- Evaluation heuristic
- How do we measure classifier quality?

hials

Testing data

Step 1

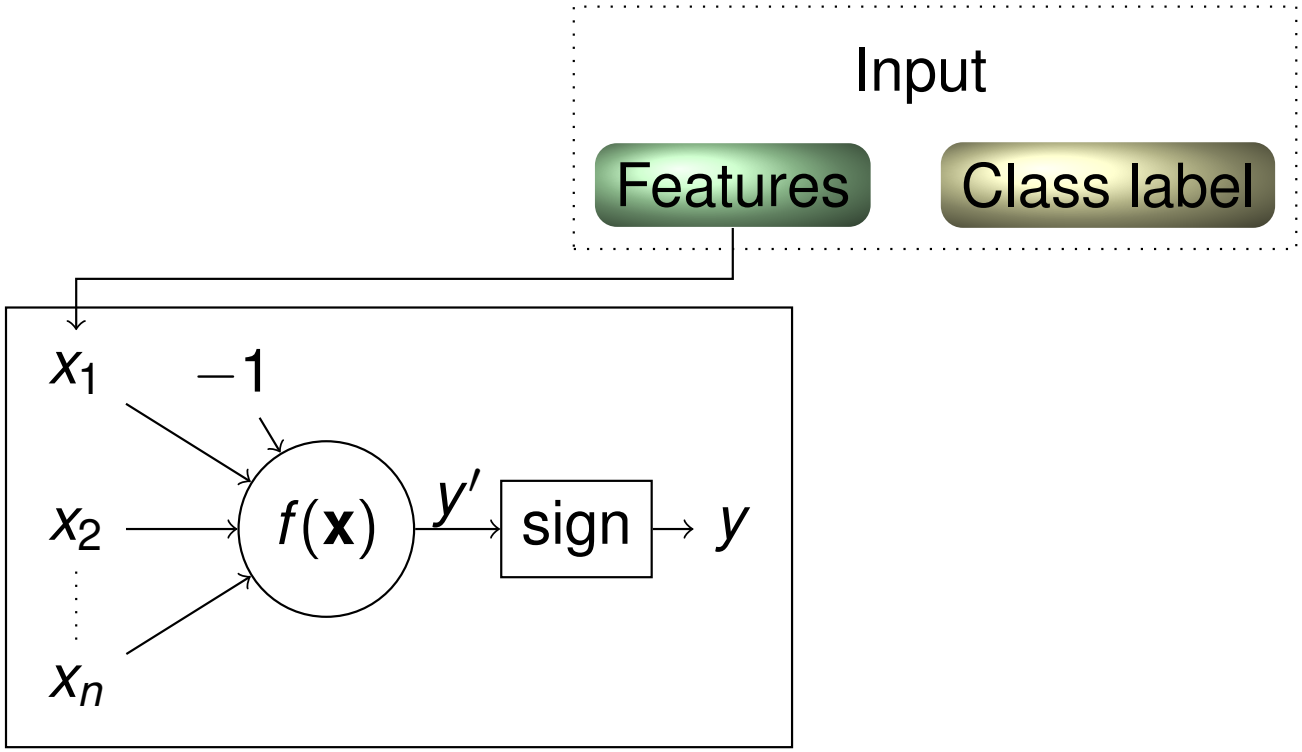
To test, we need data

1. Split the data set into two parts
   1. n_1 rows for training
   2. n_2 rows for testing
2. Make sure either set is representative
   - How?

hials

Testing the Neuron

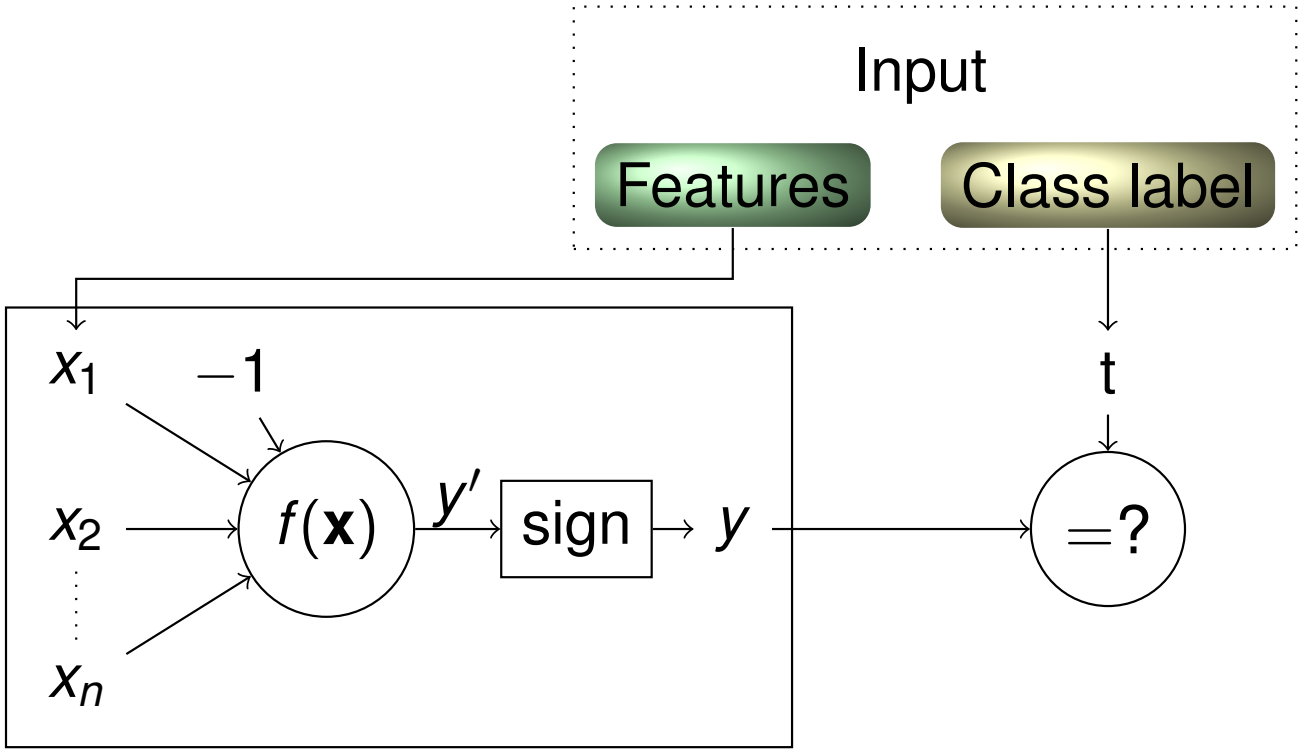
Step 2



hials

Testing the Neuron

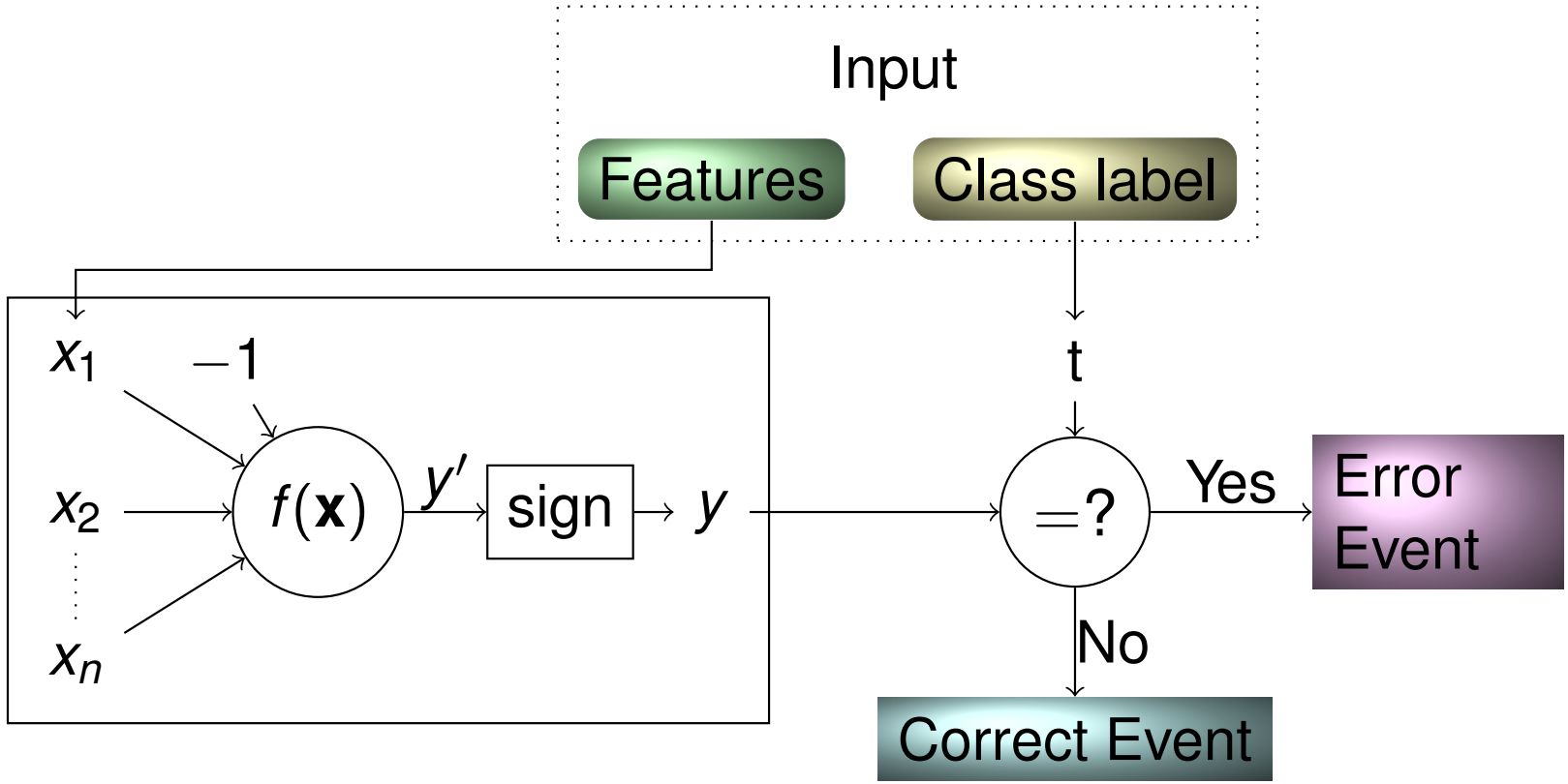
Step 2



hials

Testing the Neuron

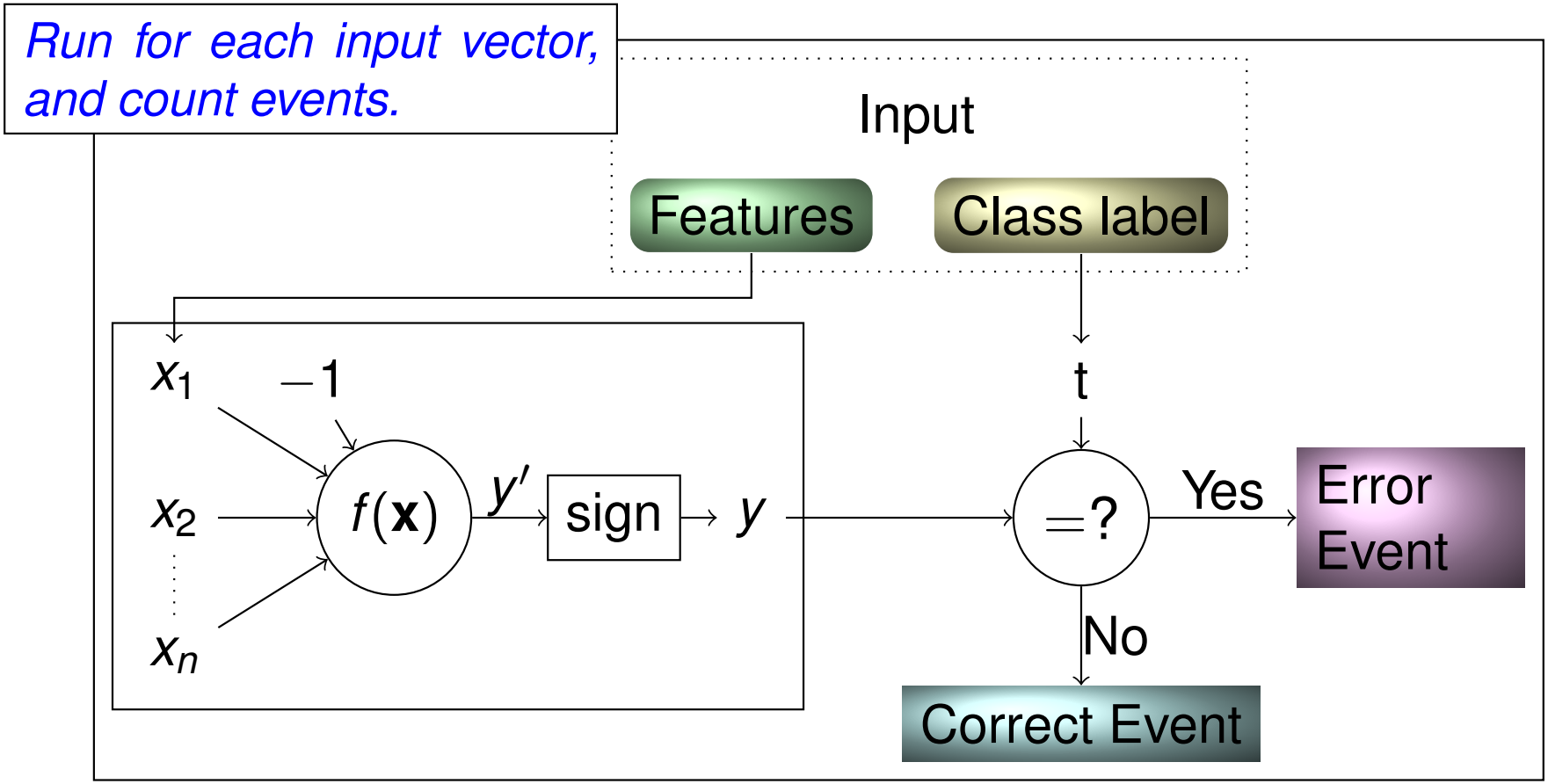
Step 2



hials

Testing the Neuron

Step 2



hials

Error Rate

1. Count
   1. e error events
   2. e correct classifications
2. Calculate error rate: $\frac{e}{e+c}$

hials

Outline

1 Error Estimation

2 Statistical Estimation

3 Evaluation Heuristic

4 Summary

hials

Error probability

1. What is the probability of an error?
   1. when recall is used
   2. on a random object in the wild?

hials

Point estimation

- Parameter: error probability p_e
- Estimator: error rate $\hat{p}_e$ (stochastic variable)
- Estimate: observation of the estimator r_e (error rate)

hials

Interval estimation

- Upper and lower estimator (l, u)
- p% confidence interval
- p% probability that $p_e \in (l, u)$

hials

Outline

1 Error Estimation

2 Statistical Estimation

3 Evaluation Heuristic

4 Summary

hials

Confusion Matrix

		Prediction	
		Malign	Benign
Actual Class	Malign	True Positive	False Negative
	Benign	False Positive	True Negative

1. We can count separate error rates for
   - false positives
   - false negatives
2. and estimate separate error probabilities

hials

Confusion Matrix

		Prediction	
		Malign	Benign
Actual Class	Malign	True Positive	False Negative
	Benign	False Positive	True Negative

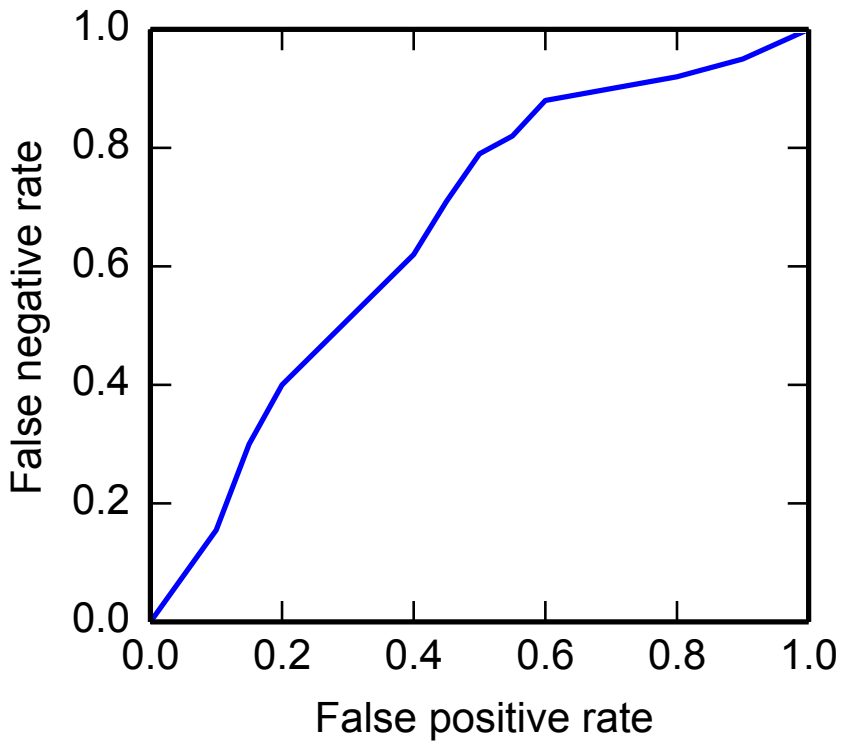
1. We can count separate error rates for
   - false positives
   - false negatives
2. and estimate separate error probabilities

hials

The ROC curve

Variable threshold

- We can vary w_0
- Receiver operating characteristic (ROC)

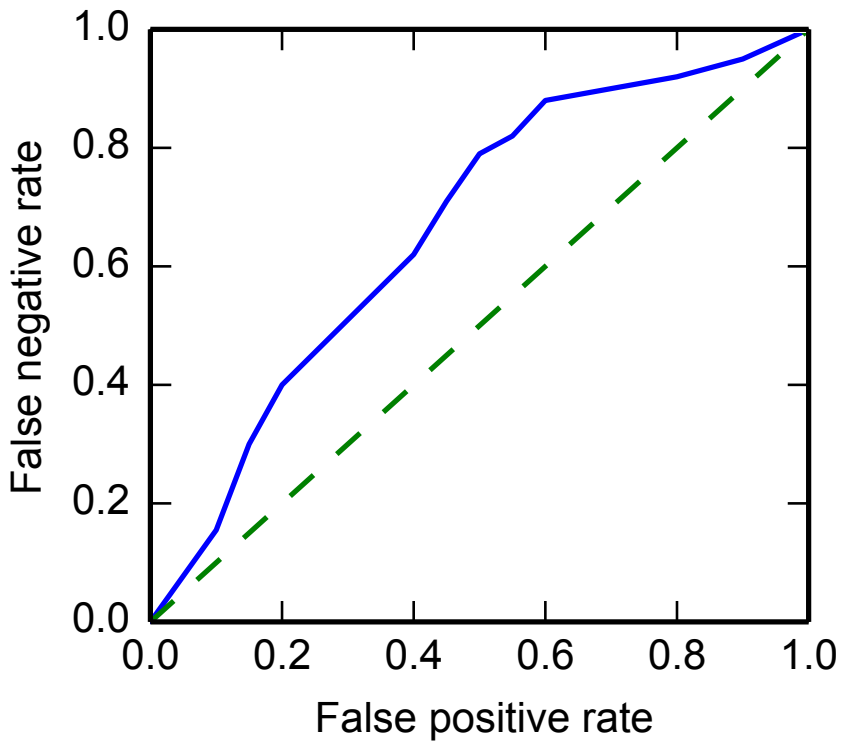


hials

The ROC curve

Variable threshold

- We can vary w_0
- Receiver operating characteristic (ROC)

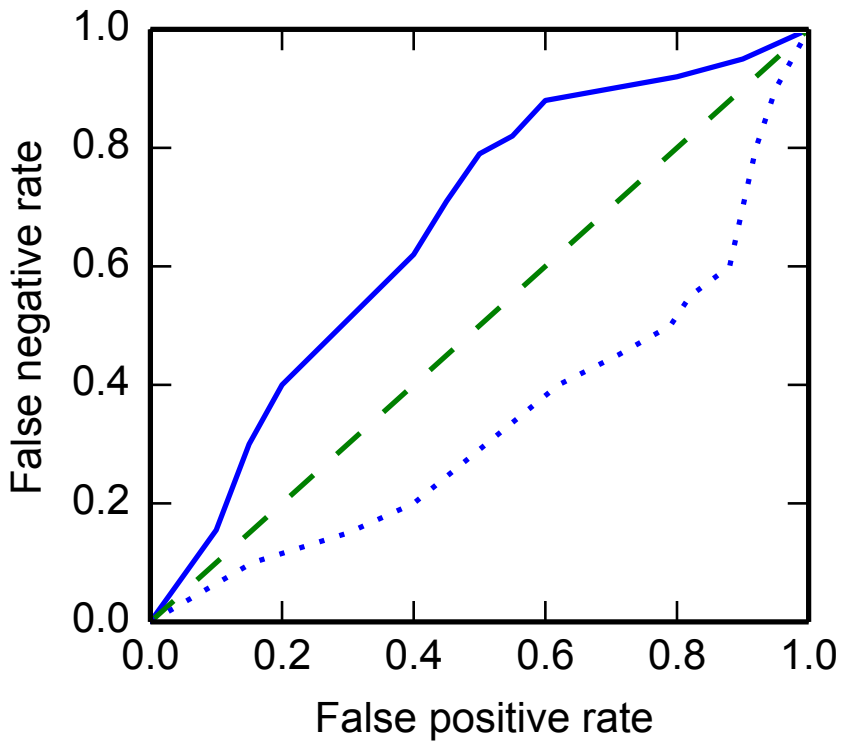


hials

The ROC curve

Variable threshold

- We can vary w_0
- Receiver operating characteristic (ROC)

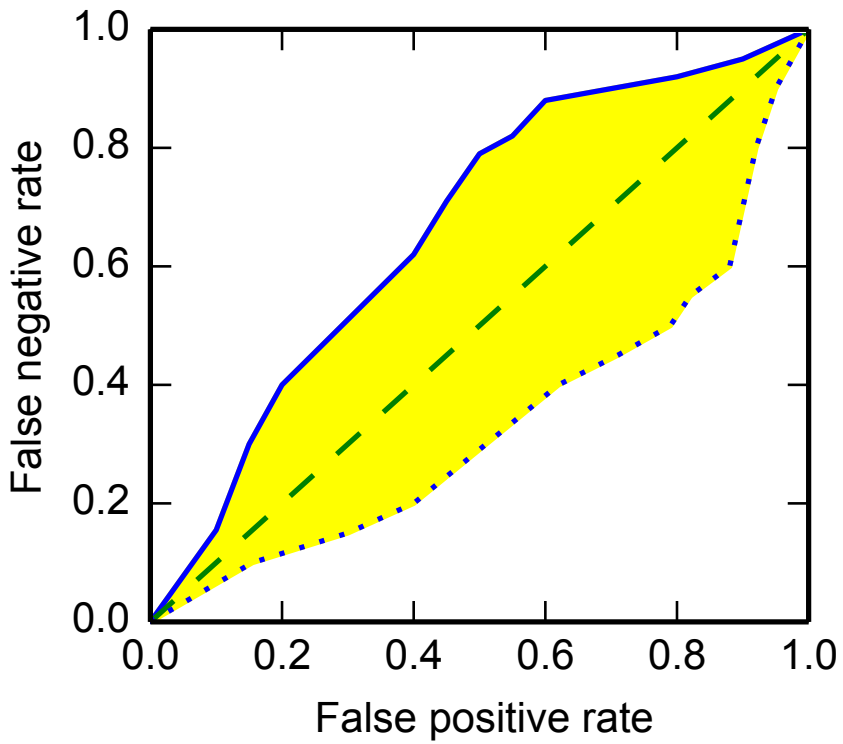


hials

The ROC curve

Variable threshold

- We can vary w_0
- Receiver operating characteristic (ROC)

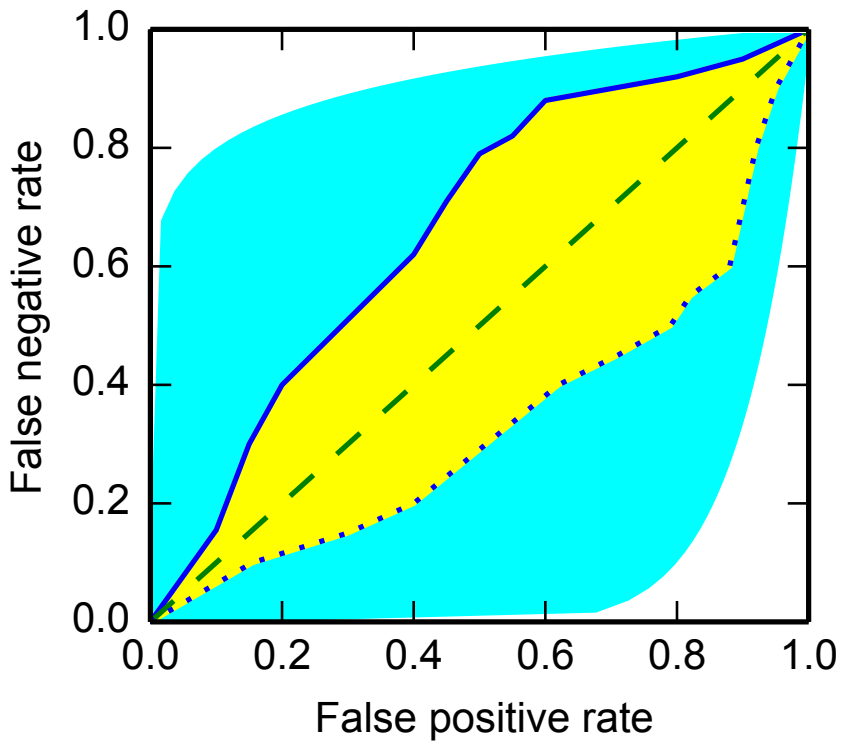


hials

The ROC curve

Variable threshold

- We can vary w_0
- Receiver operating characteristic (ROC)



hials

Outline

1 Error Estimation

2 Statistical Estimation

3 Evaluation Heuristic

4 Summary

hials

Summary

Testing

1. Use a test set independent of the training set
2. Test set with known class labels
3. Do recall, and compare to known labels

Evaluation

1. Statistical analysis of test results
2. How large test set?

hials