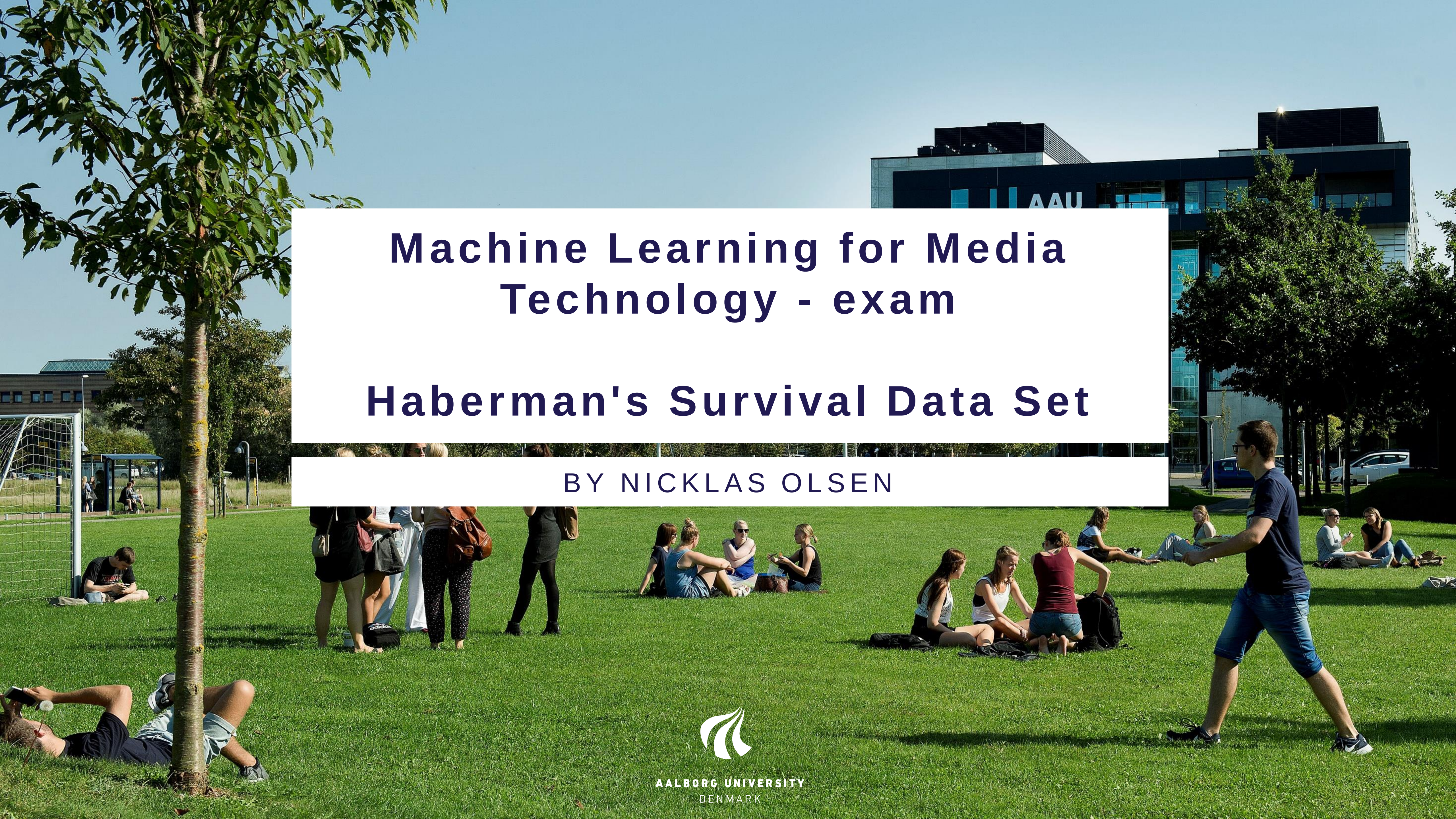


Machine Learning for Media Technology - exam

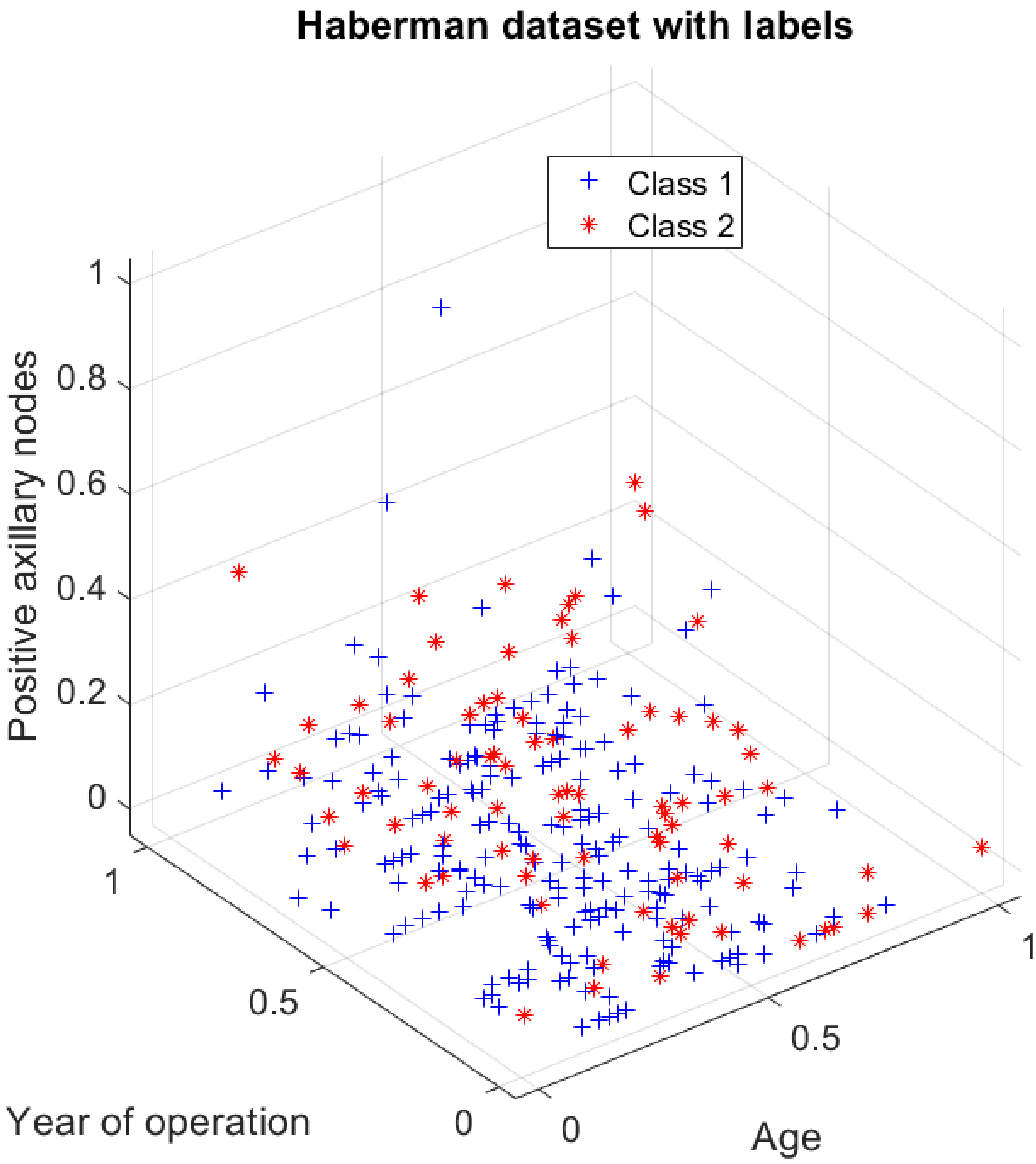
Haberman's Survival Data Set

BY NICKLAS OLSEN

AALBORG UNIVERSITY
DENMARK

DATA

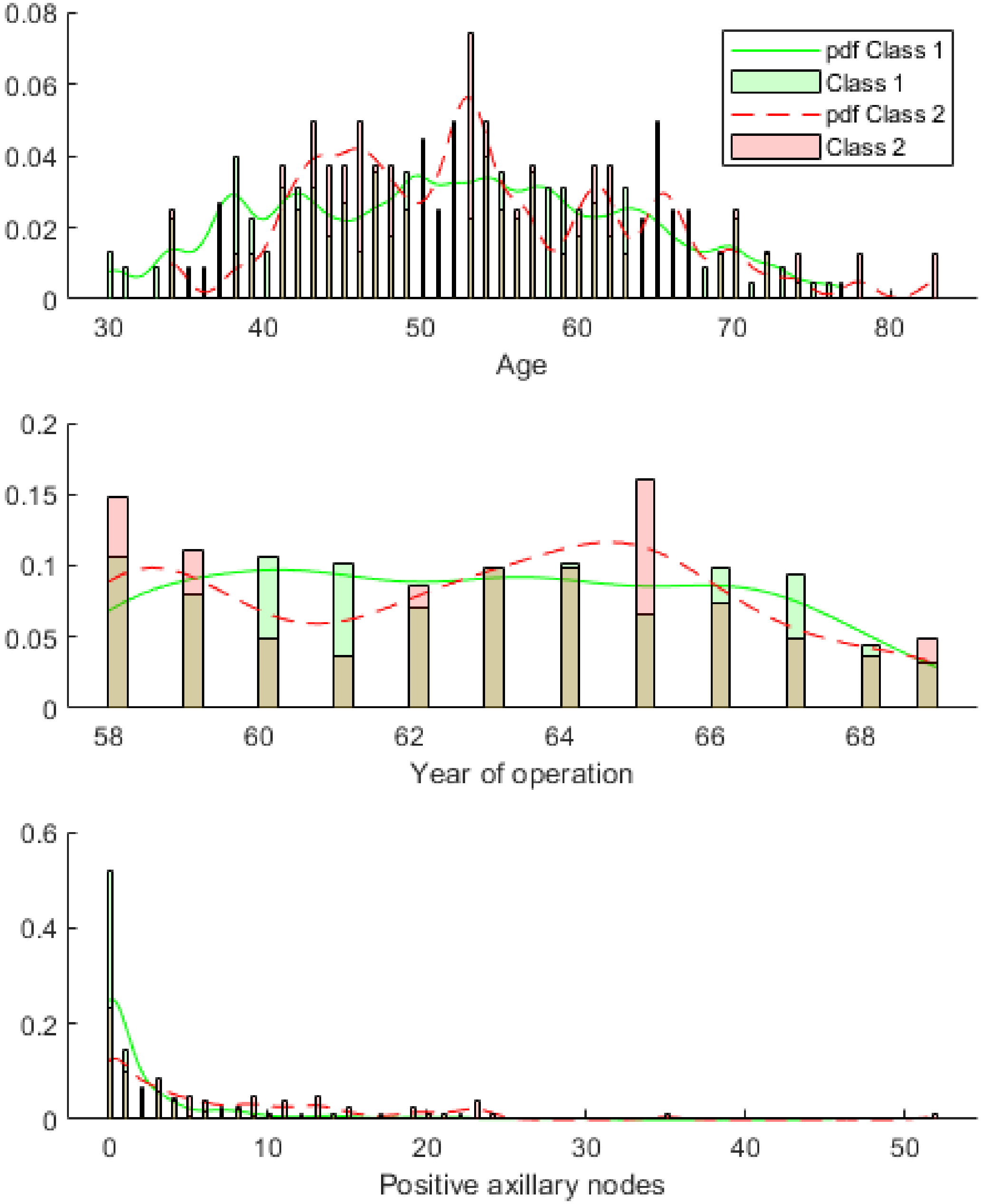
- 306 entries
- 3 Features:
  - Age
  - Year of operation
  - Positive Axillary nodes
- 2 Classes:
  1. The patient survived 5 years or longer
  2. The patient died within 5 years



AALBORG UNIVERSITY
DENMARK

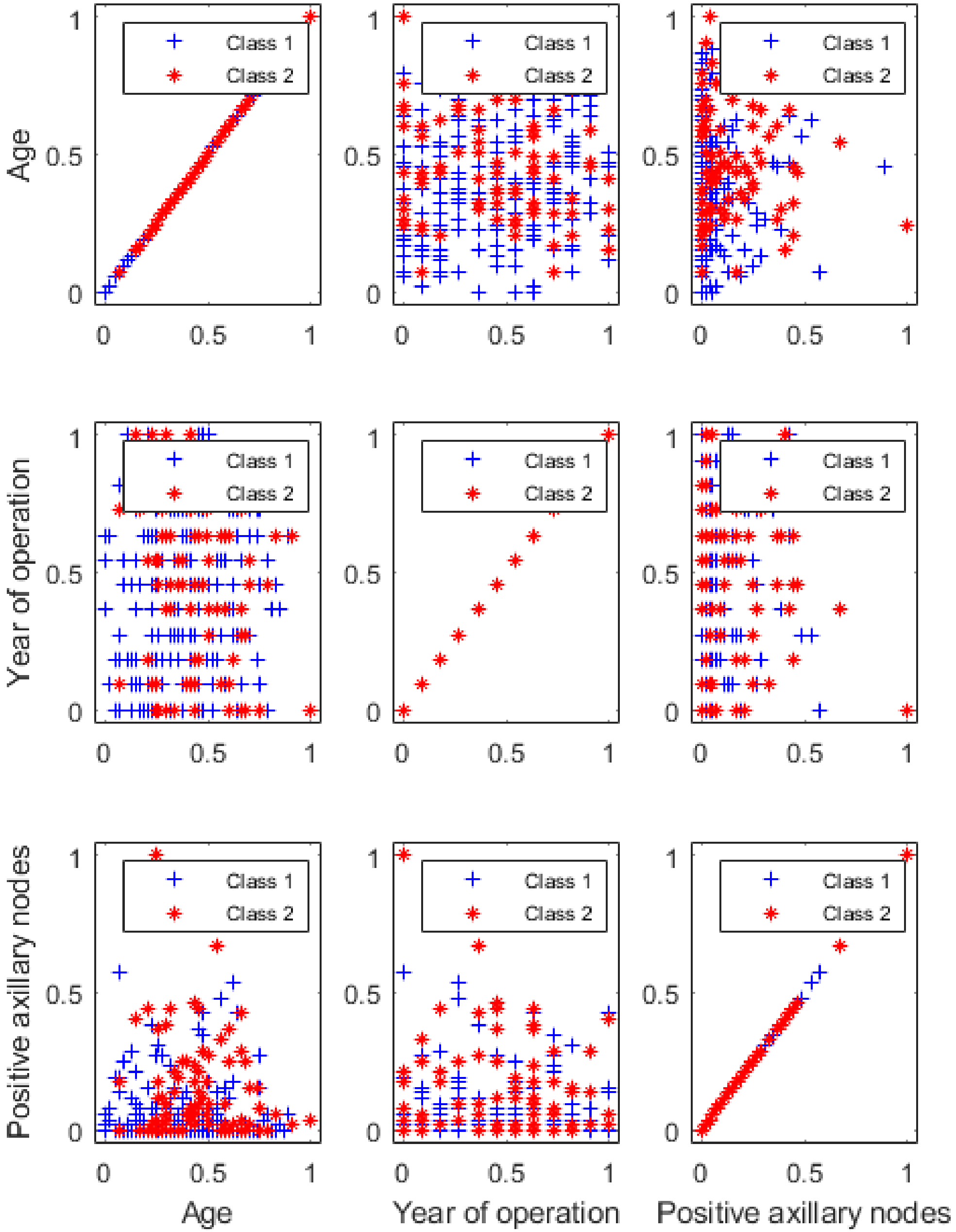
DATA

- 306 entries
- 3 Features:
  - Age
  - Year of operation
  - Positive Axillary nodes
- 2 Classes:
  1. The patient survived 5 years or longer
  2. The patient died within 5 years



DATA

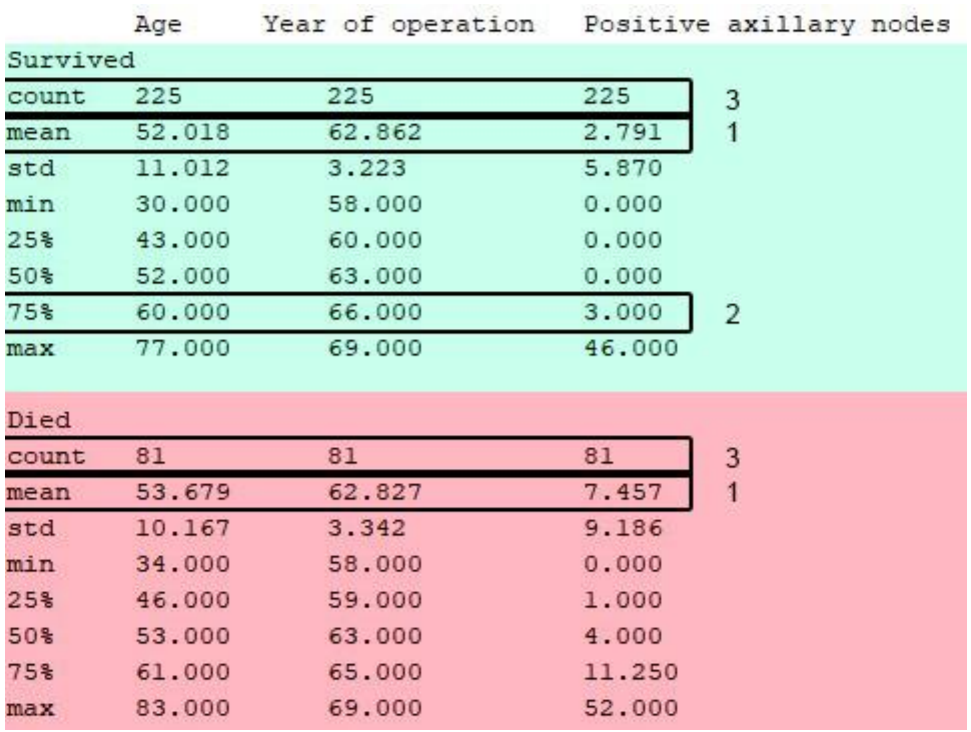
- 306 entries
- 3 Features:
  - Age
  - Year of operation
  - Positive Axillary nodes
- 2 Classes:
  1. The patient survived 5 years or longer
  2. The patient died within 5 years



AALBORG UNIVERSITY
DENMARK

DATA ANALYSIS

- Amount
- Means
- Min/Max
- Quantiles
- Distributions
  - Priors
  - Covariances

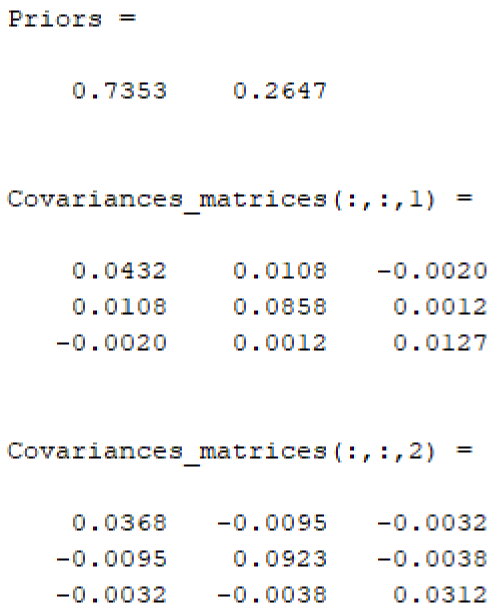


	Age	Year of operation	Positive axillary nodes	
Survived				
count	225	225	225	3
mean	52.018	62.862	2.791	1
std	11.012	3.223	5.870	
min	30.000	58.000	0.000	
25%	43.000	60.000	0.000	
50%	52.000	63.000	0.000	
75%	60.000	66.000	3.000	2
max	77.000	69.000	46.000	
Died				
count	81	81	81	3
mean	53.679	62.827	7.457	1
std	10.167	3.342	9.186	
min	34.000	58.000	0.000	
25%	46.000	59.000	1.000	
50%	53.000	63.000	4.000	
75%	61.000	65.000	11.250	
max	83.000	69.000	52.000	

AALBORG UNIVERSITY
DENMARK

DATA ANALYSIS

- Amount
- Means
- Min/Max
- Quantiles
- Distributions
  - Priors
  - Covariances



```
Priors =

    0.7353    0.2647


Covariances_matrices(:,:,1) =

    0.0432    0.0108   -0.0020
    0.0108    0.0858    0.0012
   -0.0020    0.0012    0.0127


Covariances_matrices(:,:,2) =

    0.0368   -0.0095   -0.0032
   -0.0095    0.0923   -0.0038
   -0.0032   -0.0038    0.0312
```

Feature selection

- Inter/intra class distance
  - For each possible subset permutation

$$J_{INTER/INTRA} = trace\left(\mathbf{S}_w^{-1}\mathbf{S}_b\right)$$

[Age, Year of operation, Positive axillary nodes]

	[1,2,3] 0.2514	
[1,2] 0.0121	[1,3] 0.2510	[2,3] 0.2283
[1] 0.0118	[2] 0.000058	[3] 0.2283

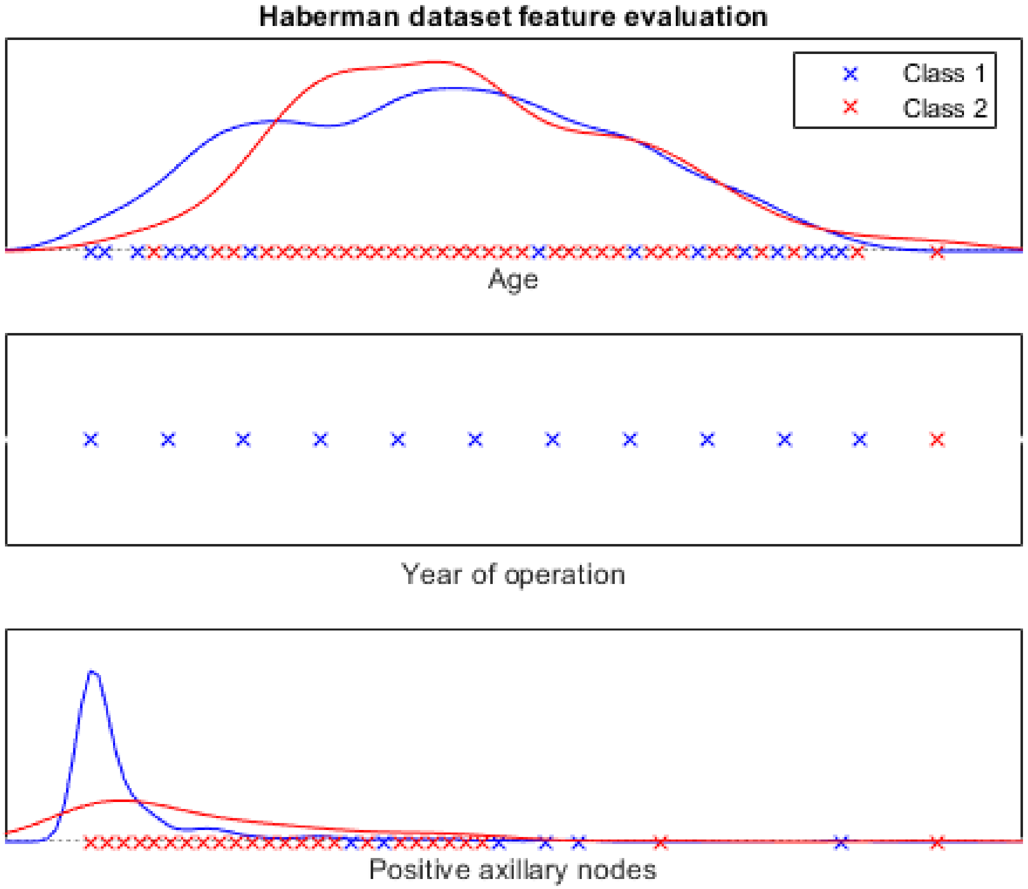
Feature selection (visual inspection)

- How do these results correlate with an visual insepection of each feature?

[1,2,3]
0.2514

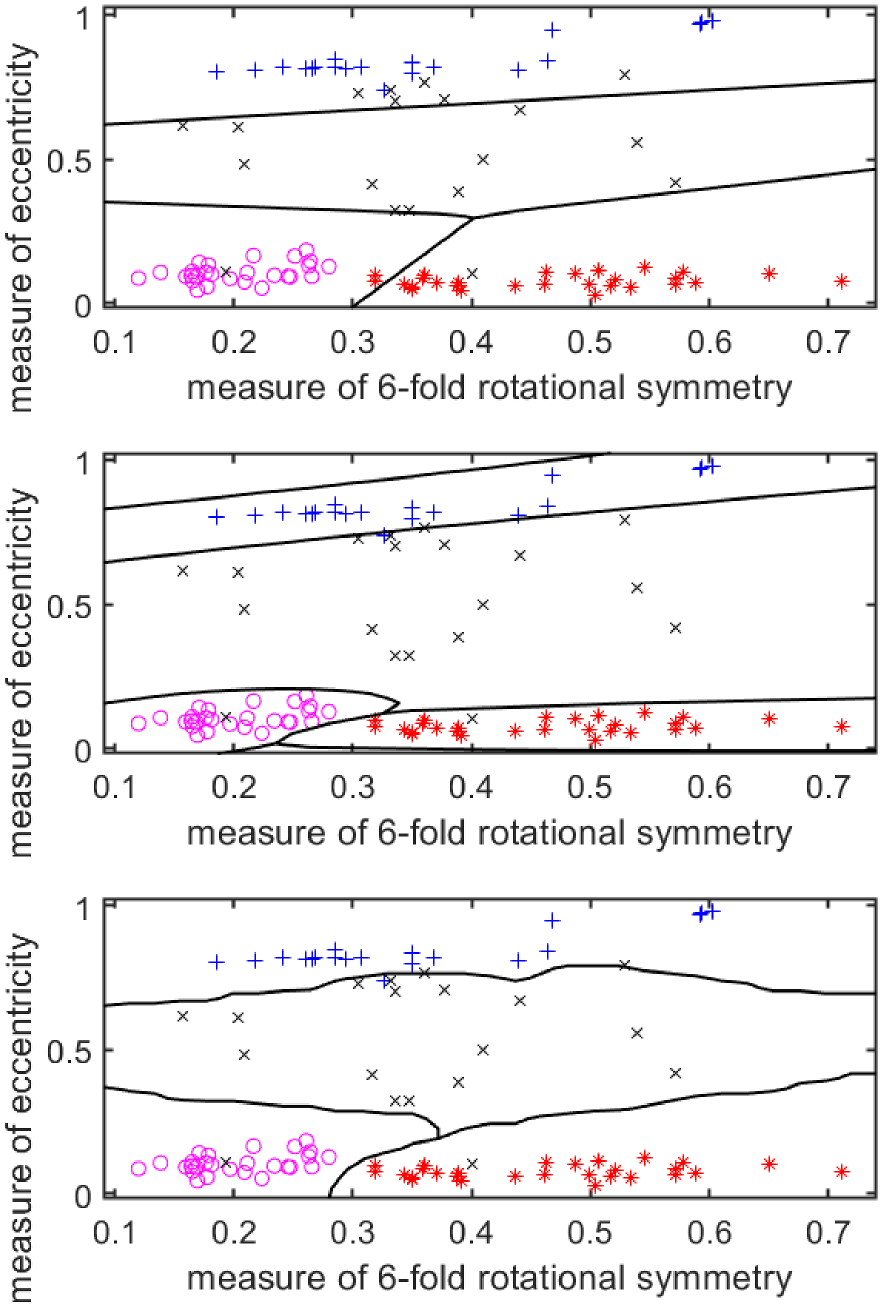
[1,2] 0.0121 [1,3] 0.2510 [2,3] 0.2283

[1] 0.0118 [2] 0.000058 [3] 0.2283



Training

- With these results
  - Hard for any classifier to classify with high precision
- Evaluate the following classifiers:
  - Linear classifier (ldc)
    - Clearly not cut out for such a task
  - Quadratic classifier (qdc)
    - Has potential
  - Nearest Neighbor Classifier (knnc)
    - Has potential

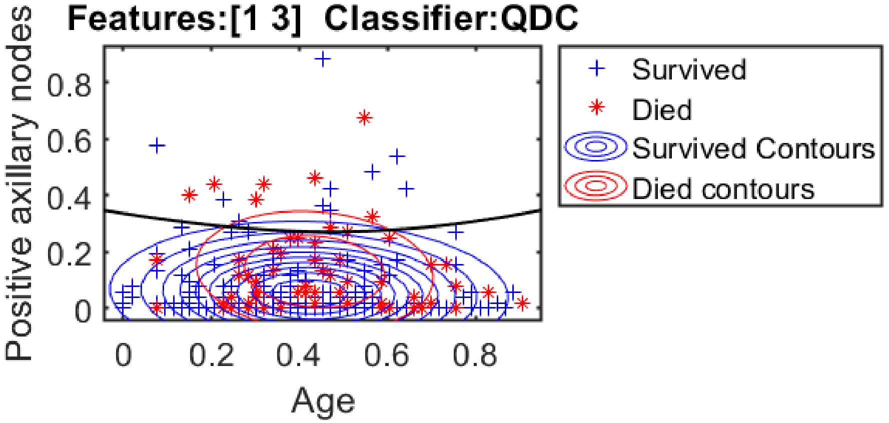


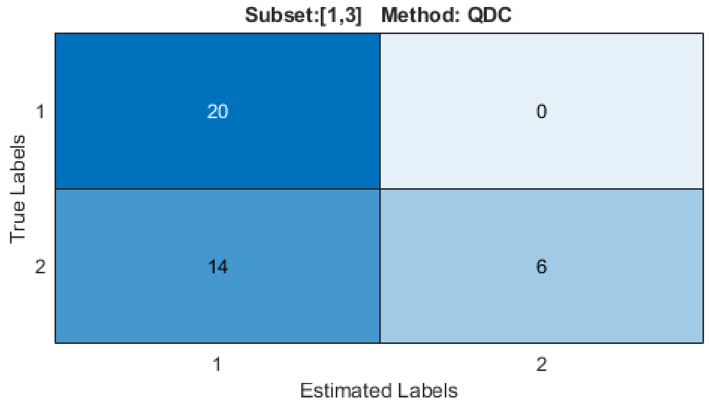
Results

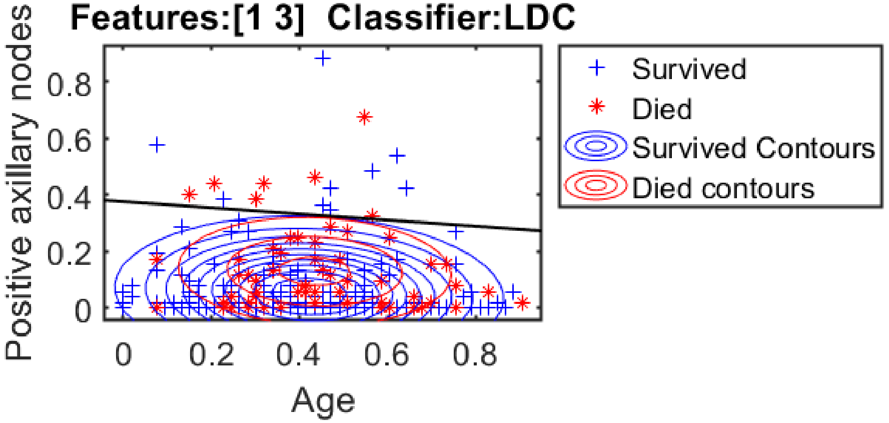
- Best subset
  - Cross validation:
    - qdc: [3] ⟶ 0.236842
    - ldc: [3] ⟶ 0.233083
    - knnc:[3] ⟶ 0.221805
  - Performance estimation (testc):
    - qdc: [3] ⟶ 0.35
    - ldc: [3] ⟶ 0.375
    - knnc: [1,3] ⟶0.45

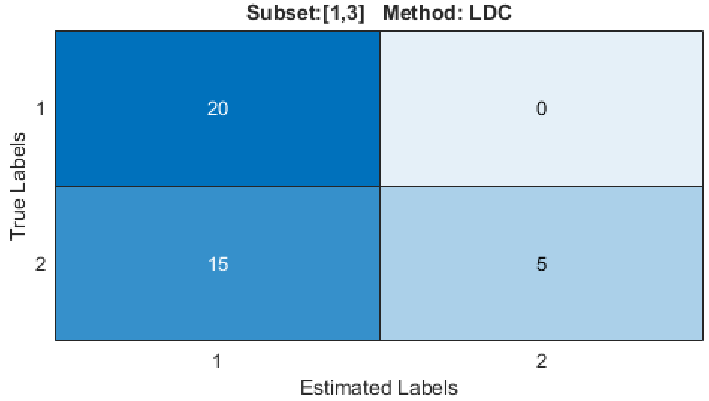
[1,3]

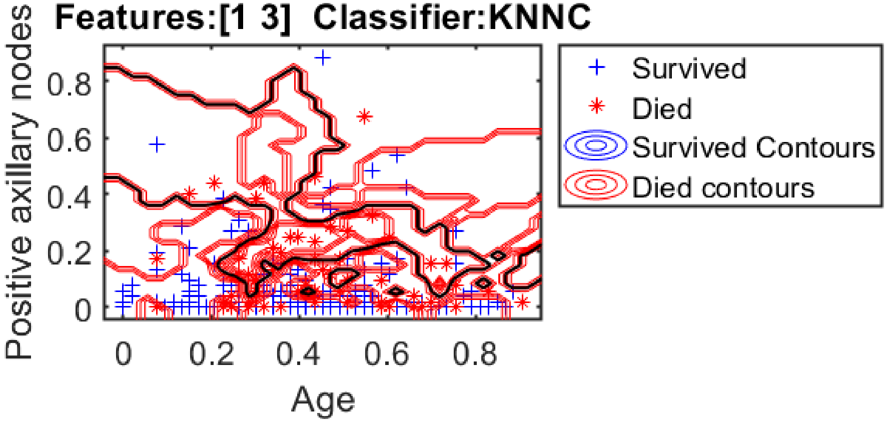
- ldc
  - Decision boundary plot
  - Confusion matrix
  - ROC plot
- qdc
  - Decision boundary plot
  - Confusion matrix
  - ROC plot
- knnc
  - Decision boundary plot
  - Confusion matrix
  - ROC plot

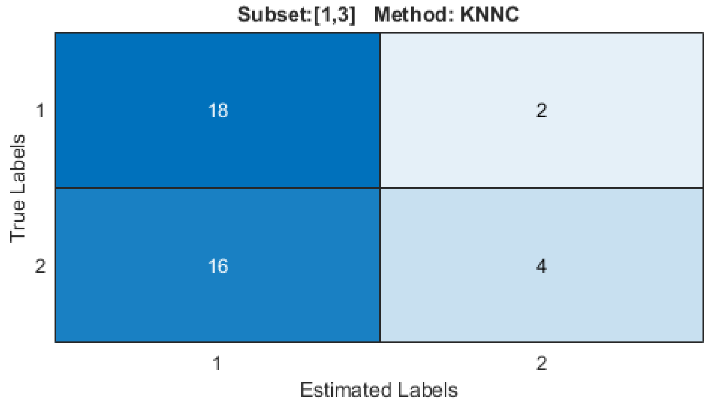






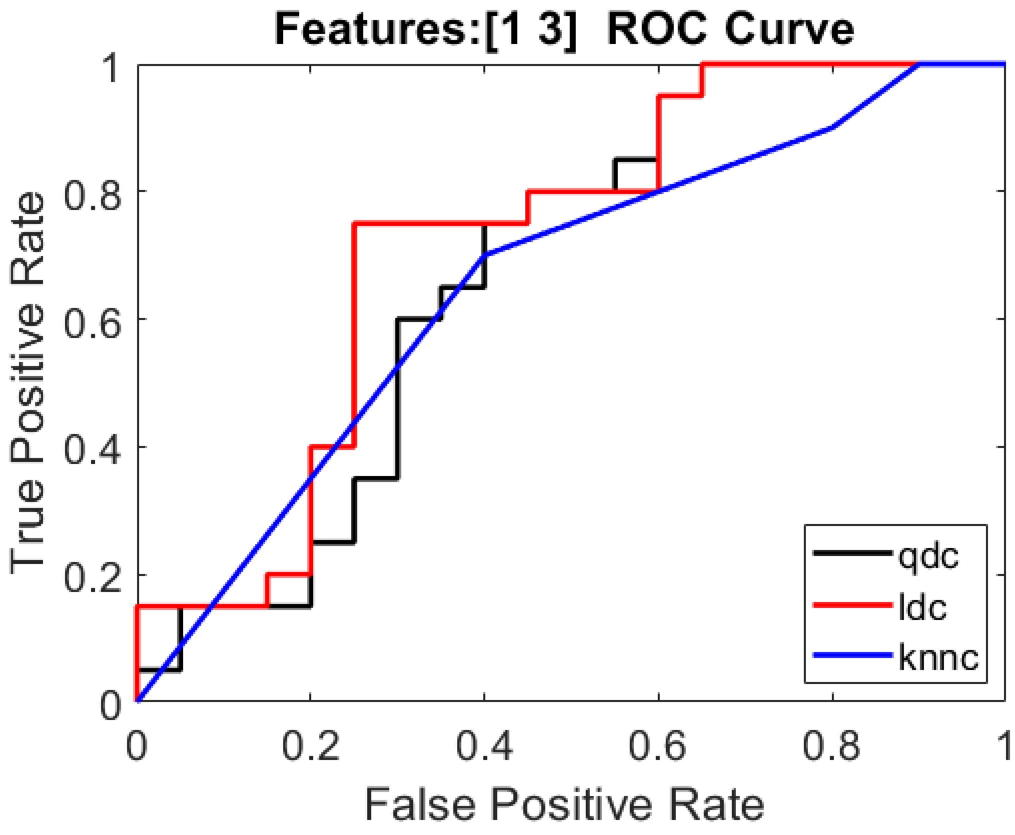






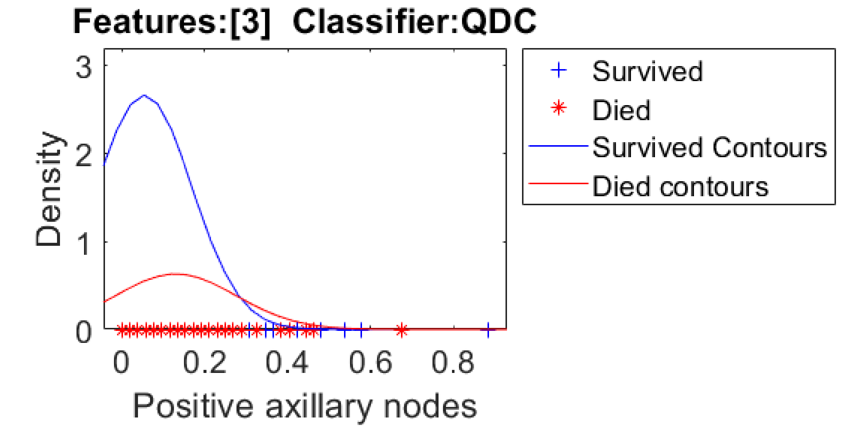
[1,3]

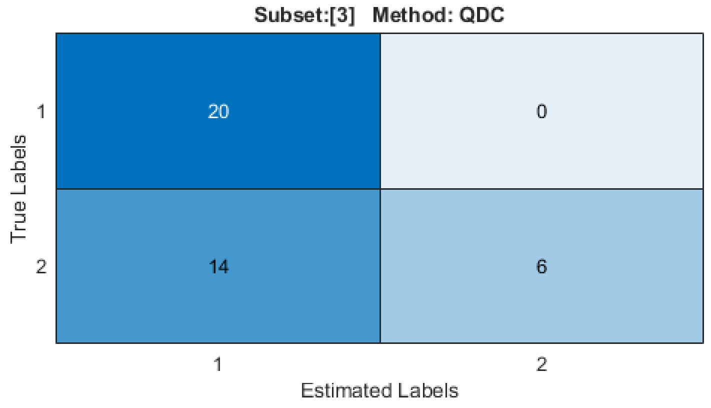
- ldc
  - Decision boundary plot
  - Confusion matrix
  - ROC plot
- qdc
  - Decision boundary plot
  - Confusion matrix
  - ROC plot
- knnc
  - Decision boundary plot
  - Confusion matrix
  - ROC plot

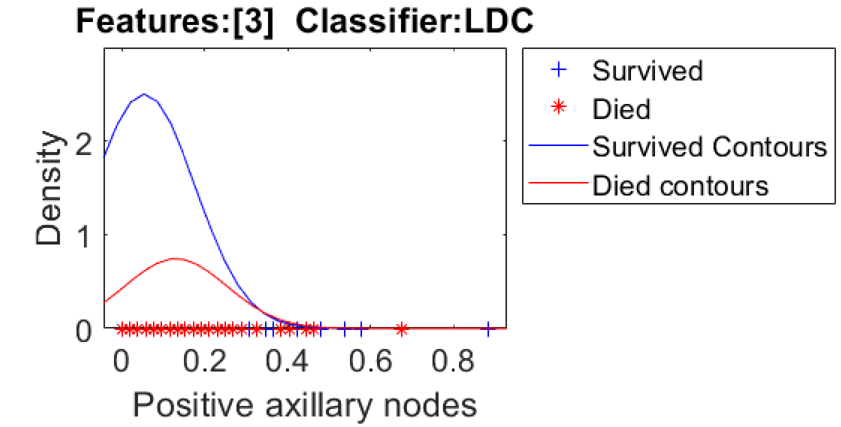


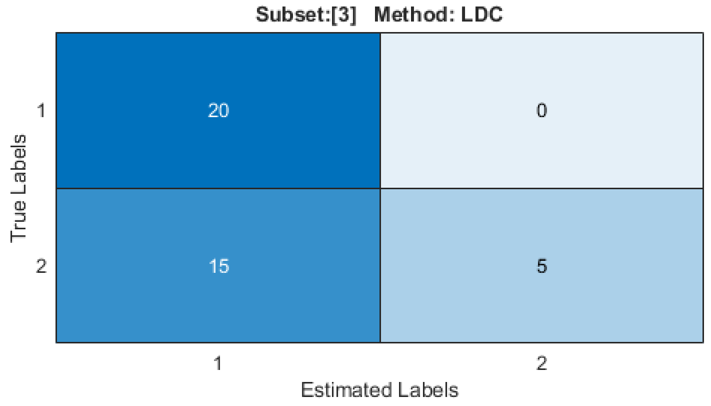
[3]

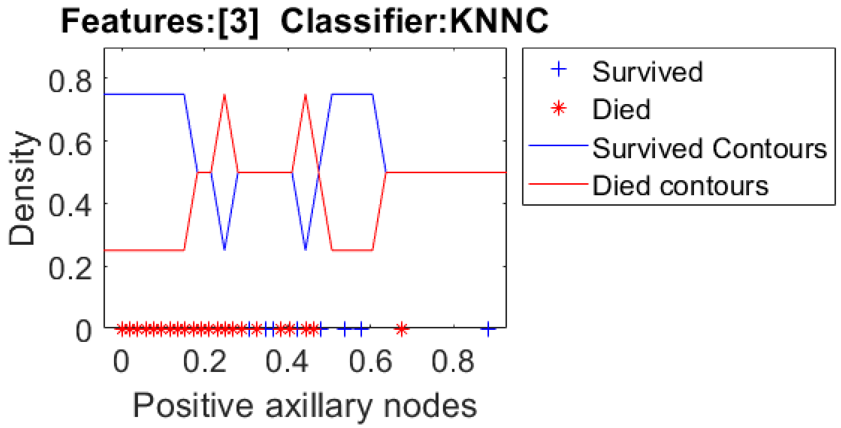
- ldc
  - Contour plot
  - Confusion matrix
  - ROC plot
- qdc
  - Contour plot
  - Confusion matrix
  - ROC plot
- knnc
  - Contour plot
  - Confusion matrix
  - ROC plot

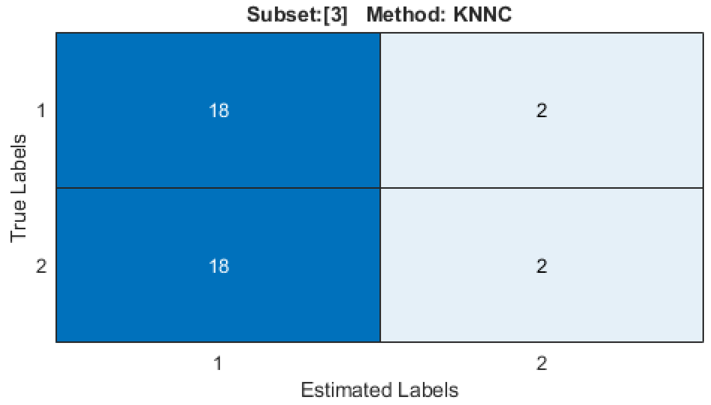

Features:[3] Classifier:QDC


Subset:[3] Method: QDC


Features:[3] Classifier:LDC

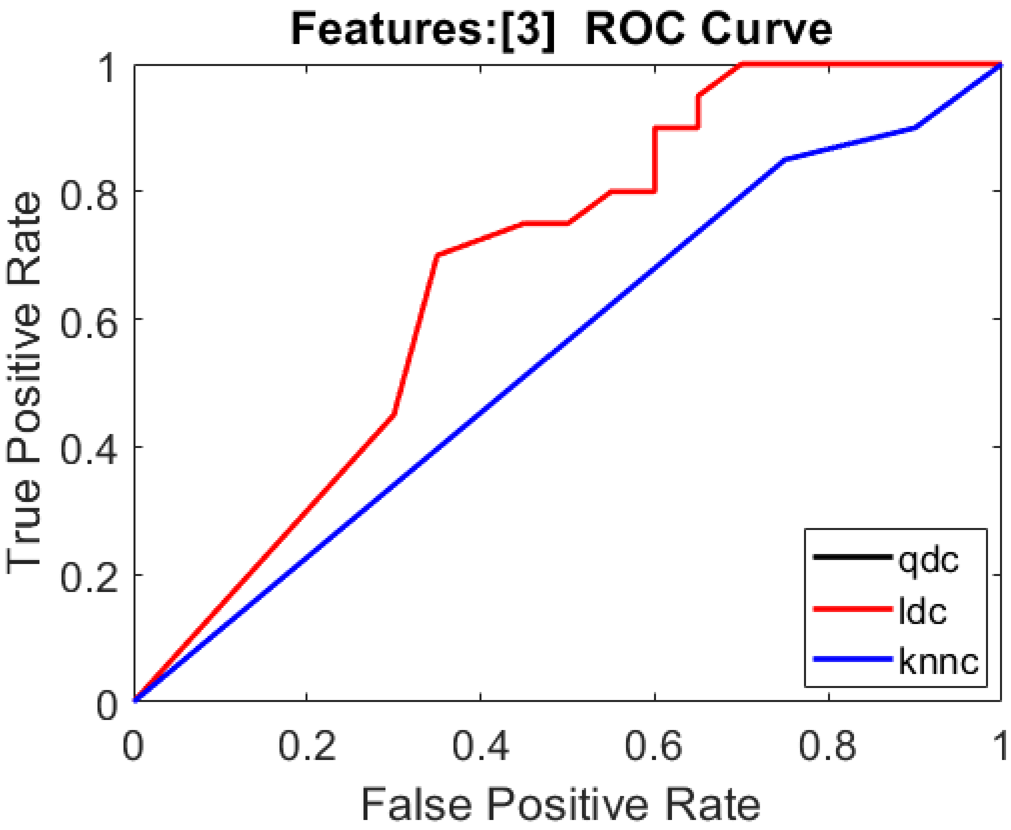

Subset:[3] Method: LDC


Features:[3] Classifier:KNNC


Subset:[3] Method: KNNC

[3]

- ldc
  - Contour plot
  - Confusion matrix
  - ROC plot
- qdc
  - Contour plot
  - Confusion matrix
  - ROC plot
- knnc
  - Contour plot
  - Confusion matrix
  - ROC plot



AALBORG UNIVERSITY
DENMARK