Rozpoznávanie obrazcov

KNN

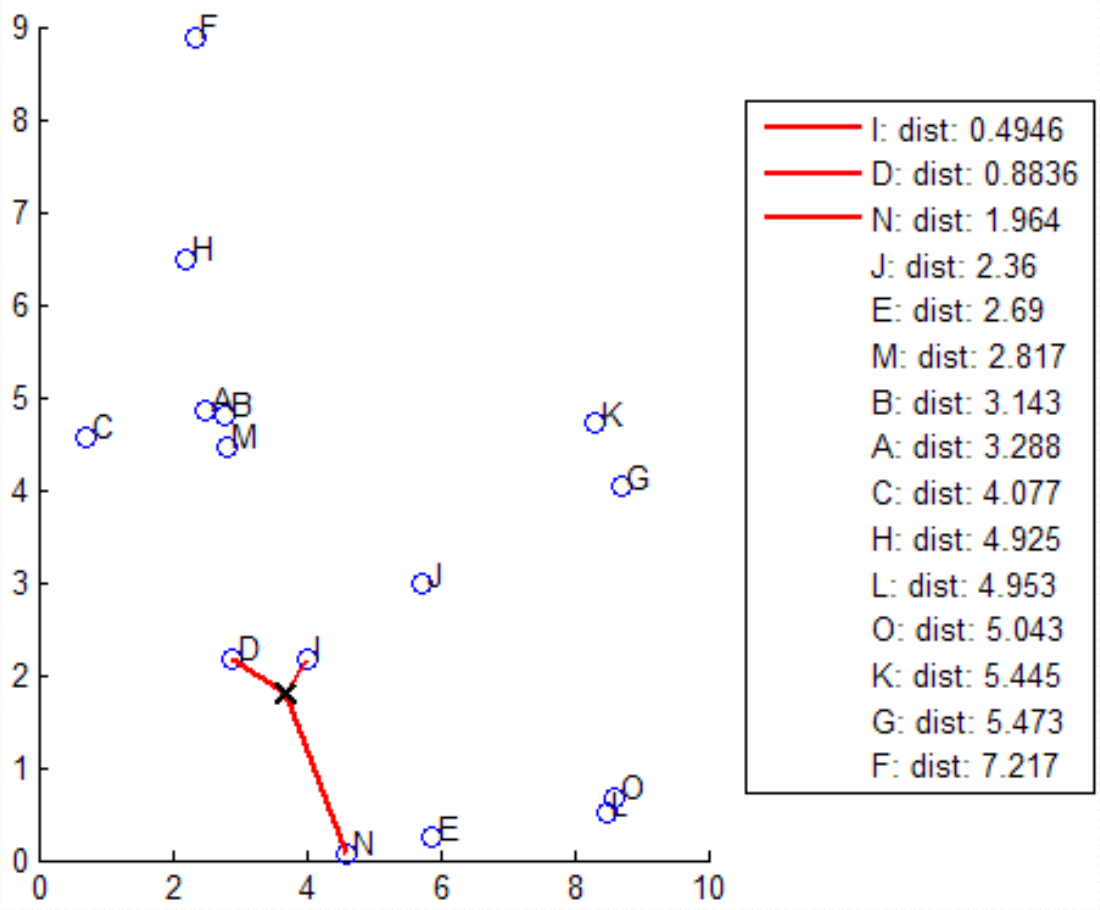
Rozhodovacie stromy

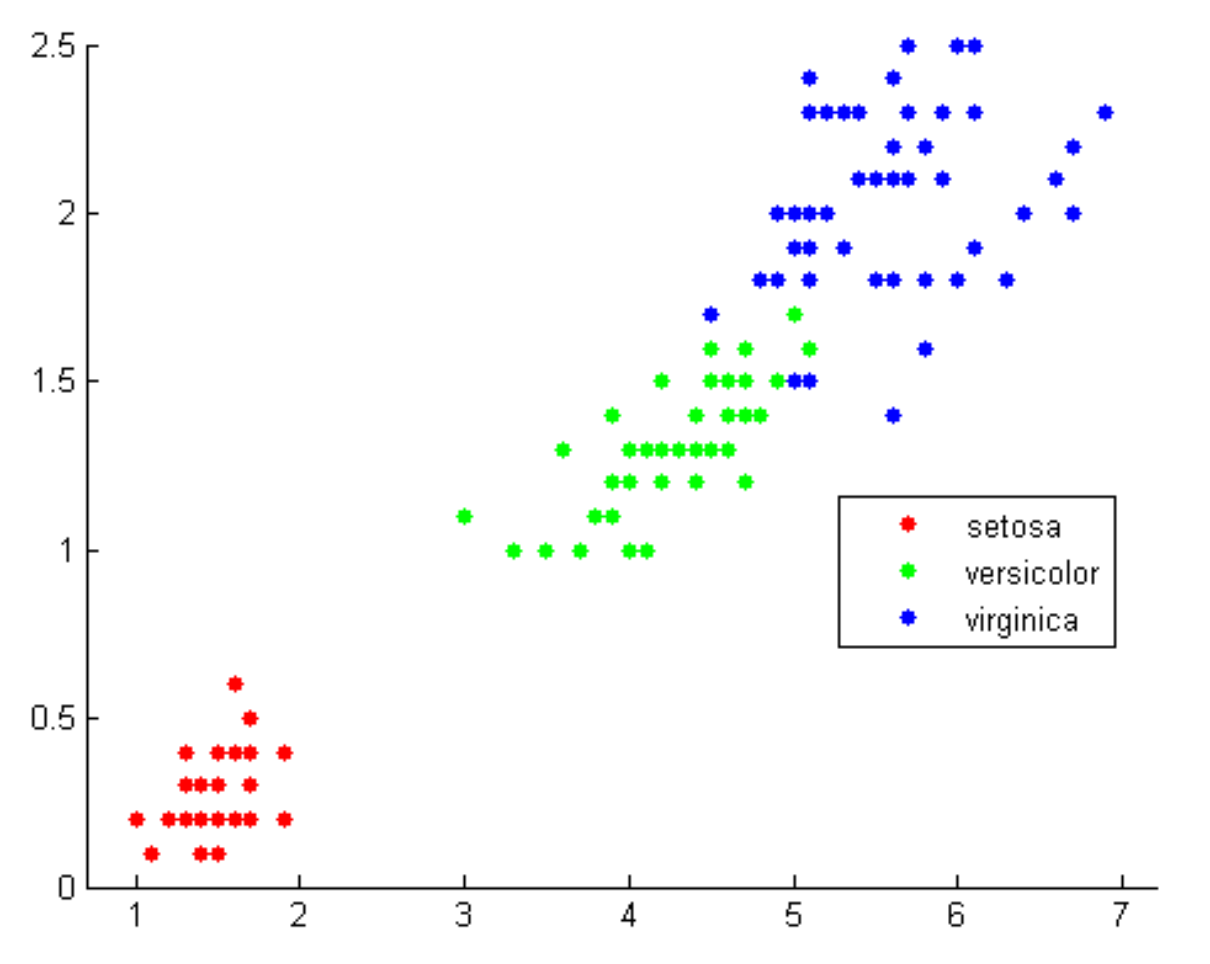
Bayesovský klasifikátor

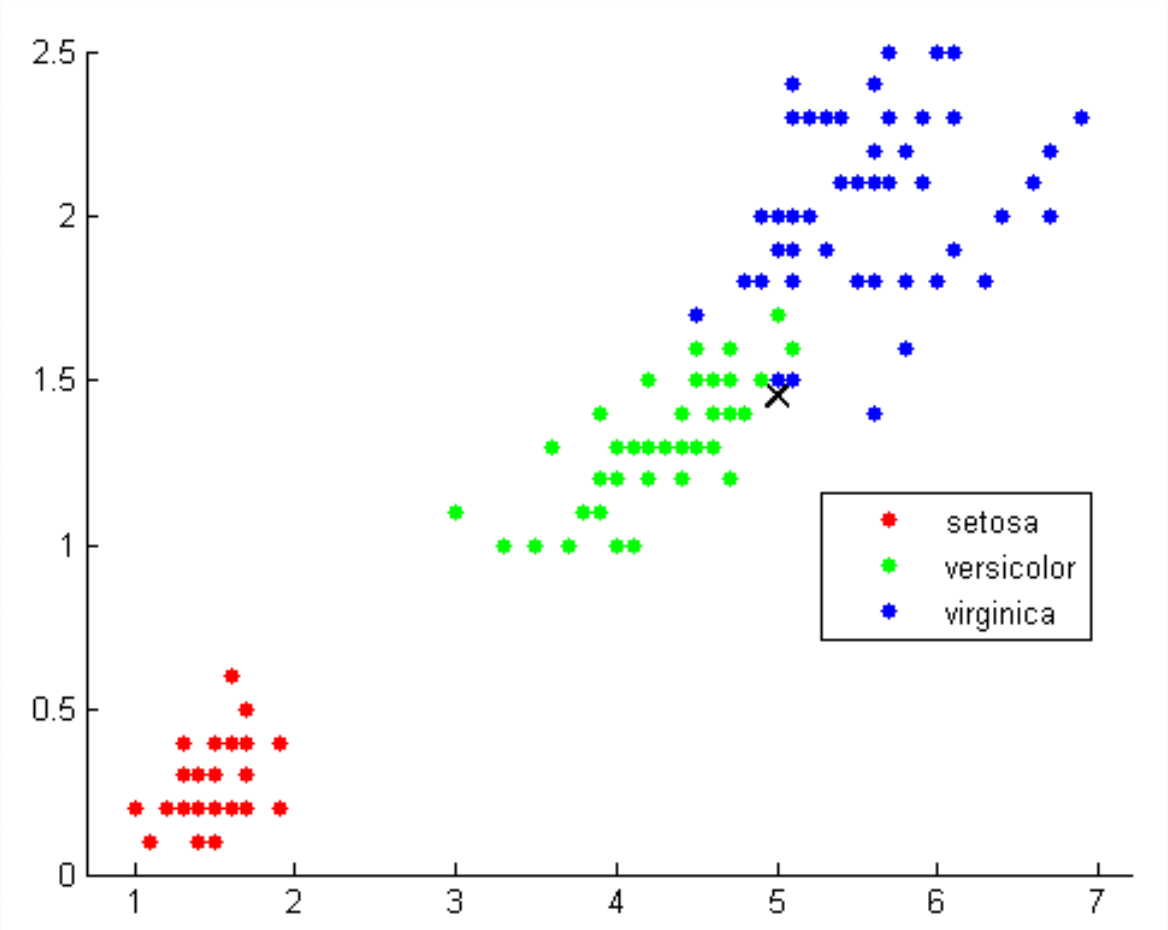
31.3.2014

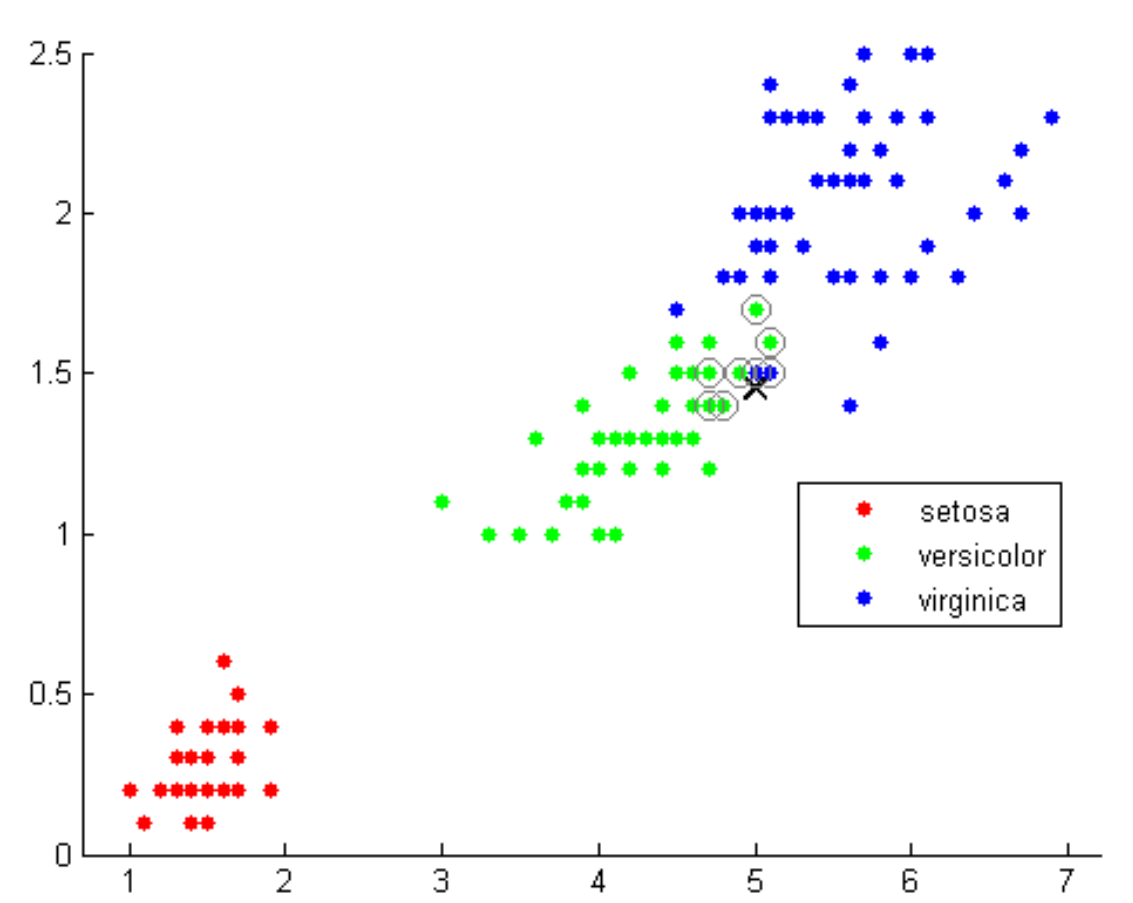
KNN

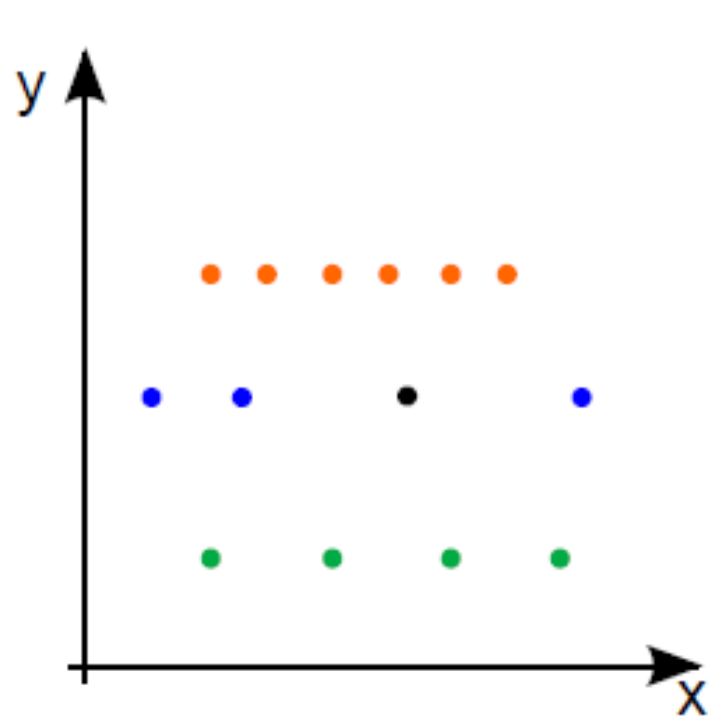
- k-nearest neighbor

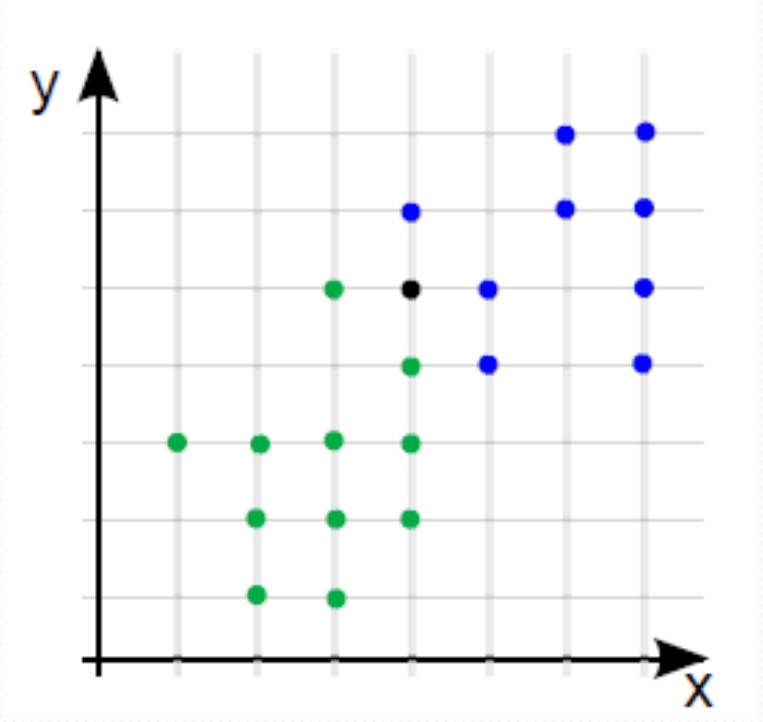


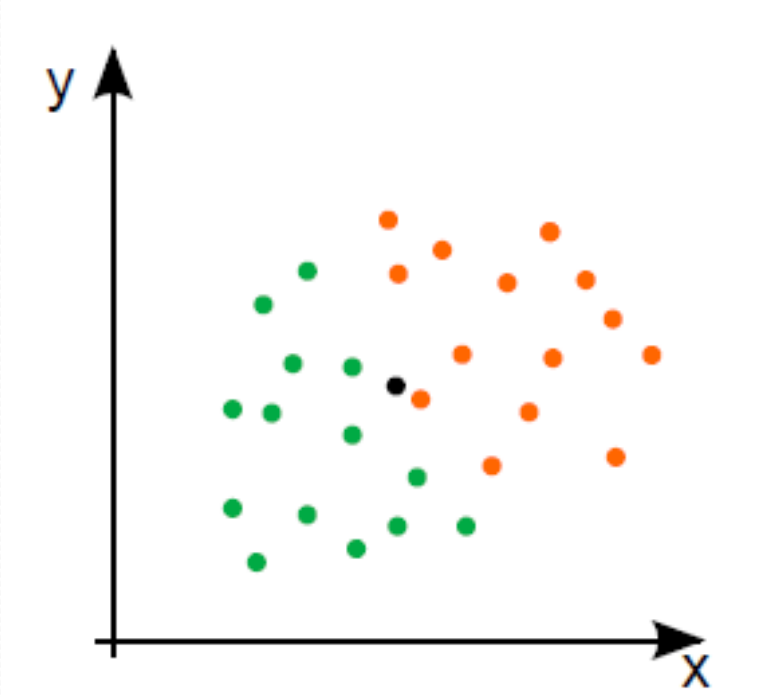

2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7
setosa
versicolor
virginica


2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7
setosa
versicolor
virginica


2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7
setosa
versicolor
virginica


y
x


y
X


y
x

KNN - example

- http://www.mathworks.com/matlabcentral/fileexchange/19345-efficient-k-nearest-neighbor-search-using-jit

```
R=randn(100,2);
Q=randn(1,2);
idx=knnsearch(Q,R, 5);
plot(R(:,1),R(:,2),'b.',Q(:,1),Q(:,2),'ro',R(idx,1),R(idx,2),'gx');
```

KNN - example

```
R=rand(100,2);
Q=[0 0];
K=10;
idx=knnsearch(Q,R,10);
r=max(sqrt(sum(R(idx,:).^2,2)));
theta=0:0.01:pi/2;
x=r*cos(theta);
y=r*sin(theta);
plot(R(:,1),R(:,2),'b.',Q(:,1),Q(:,2),'co',R(idx,1),R(idx,2),'gx'
   ,x,y,'r-','linewidth',2);
```

KNN

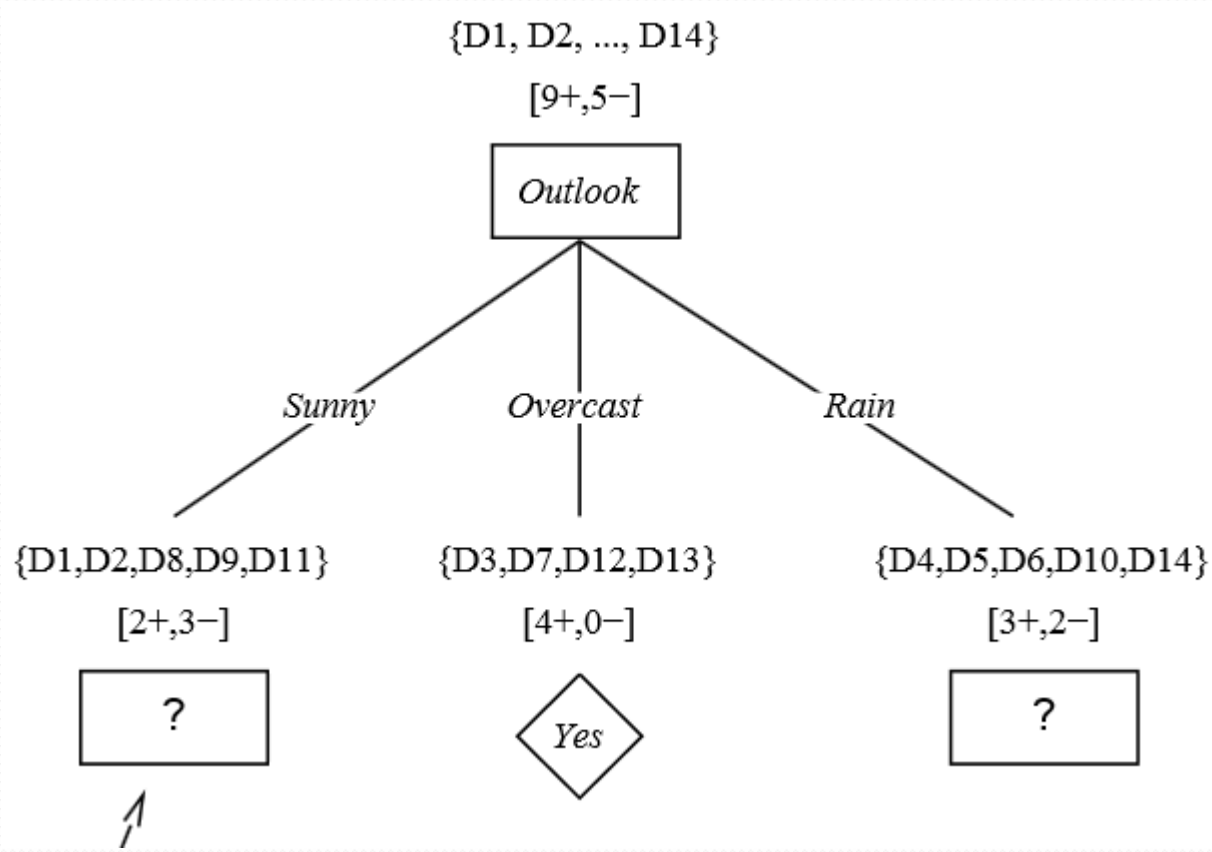
- Statistic toolbox
  - http://www.mathworks.com/help/stats/classificationknnclass.html
  - http://www.mathworks.com/help/stats/classificationknn.fit.html

Rozhodovacie stromy

- Rozhodnutie o zaradení triedy a postup, ako sme k nemu dospeli, sú usporiadané do stromovej štruktúry
- uzly
  - opisujú testy hodnôt jednotlivých príznakov
  - z uzlov vychádza toľko vetiev, koľko rôznych hodnôt test nadobúda (väčšinou binárne)
- listy
  - klasifikačné triedy

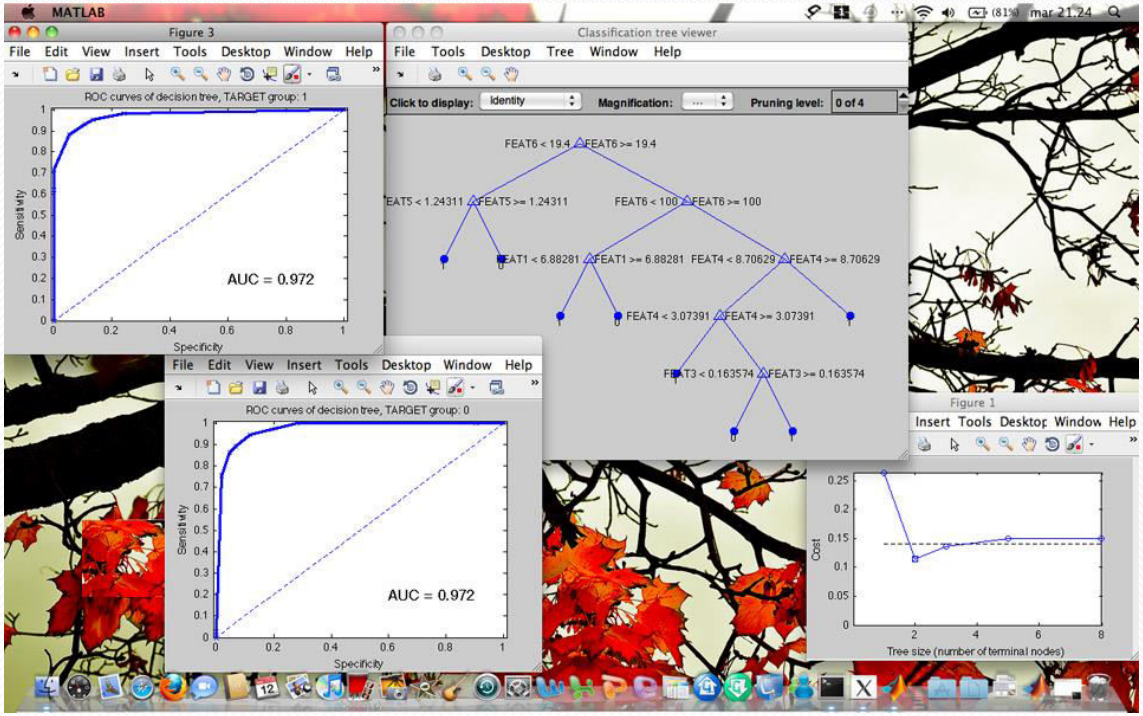
- 1. vezmi všetky nepoužité príznaky, ohodnoť
- 2. do uzlu vezmi príznak s najlepším ohodnotením
- 3. pre každú hodnotu príznaku vytvor podmnožinu dát

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No


{D1, D2, ..., D14}
[9+,5−]
Outlook
Sunny
Overcast
Rain
{D1,D2,D8,D9,D11}
[2+,3−]
?
{D3,D7,D12,D13}
[4+,0−]
Yes
{D4,D5,D6,D10,D14}
[3+,2−]
?

Rozhodovacie stromy

- http://www.mathworks.com/matlabcentral/fileexchange/26326-decision-trees-and-predictive-models-with-cross-validation-and-roc-analysis-plot



Bayesovksý klasifikátor

- Objekt je priradený do triedy, ktorá je najpravdepodobnejšia vzhľadom na príznaky
- Naivný Bayesovský klasifikátor
  - Atribúty sa navzájom neovplyvňujú

Statistical Pattern Recognition Toolbox

- http://cmp.felk.cvut.cz/cmp/software/stprtool/

Demos

- help stprtool/demos
- demo_linclass - Algorithms learning linear clasifiers.
- demo_pcacomp - Image compresion using PCA.
- demo_svm - Support Vector Machines.

KNN

```
% load training data and setup 8-NN rule
trn = load('riply_trn');
model = knnrule(trn,4);
% visualize decision boundary and training data
figure; ppatterns(trn); pboundary(model);
% evaluate classifier
tst = load('riply_tst');
ypred = knnclass(tst.X,model);
cerror(ypred,tst.y)
```

Bayes classifier

```
% load input training data
trn = load('riply_trn');
inx1 = find(trn.y==1);
inx2 = find(trn.y==2);
% Estimation of class-conditional distributions by EM
bayes_model.Pclass{1} = emgmm(trn.X(:,inx1),struct('ncomp',2));
bayes_model.Pclass{2} = emgmm(trn.X(:,inx2),struct('ncomp',2));
% Estimation of priors
n1 = length(inx1); n2 = length(inx2);
bayes_model.Prior = [n1 n2]/(n1+n2);
% Evaluation on testing data
tst = load('riply_tst');
ypred = bayescls(tst.X,bayes_model);
cerror(ypred,tst.y)
% Visualization
figure; hold on; ppatterns(trn);
bayes_model.fun = 'bayescls';
pboundary(bayes_model);
% Penalization for don't know decision
reject_model = bayes_model;
reject_model.eps = 0.1;
% Vislualization of rejet-option rule
pboundary(reject_model,struct('line_style', 'k--'));
```