

# Getting Started with your Project for Machine Learning

Imed Bouchrika  
Dept of Mathematics & Computer Science  
University of Souk Ahras  
imed@imed.ws

# Outline of the day !

- Download the training/testing dataset
- Processing with Matlab
- Classification for the Testing dataset
- Submitting your results

# Download the datasets

• Visit the following address :

• [www.imed.ws/machinelearning1](http://www.imed.ws/machinelearning1)



• Download the training + testing datasets

# Download the datasets

- Training Data :

- Categorical Attributes

- First Column : class label : { e, p }

e, x, s, y, t, a, f, c, b, k, e, c, s, s, w, w, p, w, o, p, n, n, d  
e, b, s, w, t, l, f, c, b, n, e, c, s, s, w, w, p, w, o, p, n, n, m  
p, x, y, w, t, p, f, c, n, n, e, e, s, s, w, w, p, w, o, p, k, s, u  
e, x, s, g, f, n, f, w, b, k, t, e, s, s, w, w, p, w, o, e, n, a, g  
e, b, s, w, t, a, f, c, b, g, e, c, s, s, w, w, p, w, o, p, k, n, m  
e, o, y, w, t, l, f, c, b, n, e, c, s, s, w, w, p, w, o, p, n, s, m  
p, x, y, w, t, p, f, c, n, p, e, e, s, s, w, w, p, w, o, p, k, v, g  
e, b, s, y, t, a, f, c, b, g, e, c, s, s, w, w, p, w, o, p, k, s, m  
e, x, y, y, t, a, f, c, b, n, e, c, s, s, w, w, p, w, o, p, k, s, m  
e, o, s, y, t, a, f, c, b, w, e, c, s, s, w, w, p, w, o, p, n, s, g  
p, x, y, w, t, p, f, c, n, k, e, e, s, s, w, w, p, w, o, p, n, v, u  
e, x, f, n, f, n, f, w, b, n, t, e, s, f, w, w, p, w, o, e, k, a, g  
e, f, f, w, f, n, f, w, b, k, t, e, s, s, w, w, p, w, o, e, n, a, g  
p, x, s, n, t, p, f, c, n, n, e, e, s, s, w, w, p, w, o, p, k, s, g  
p, x, y, w, t, p, f, c, n, n, e, e, s, s, w, w, p, w, o, p, n, s, u

# Download the datasets

## ● Test Data :

● Categorical Attributes

● NO class label



x,s,n,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,k,s,u  
x,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,k,n,g  
x,y,y,t,l,f,c,b,g,e,c,s,s,w,w,p,w,o,p,n,n,g  
s,f,g,f,n,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,y,u  
b,s,y,t,a,f,c,b,k,e,c,s,s,w,w,p,w,o,p,n,s,m  
f,s,w,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,n,v,g  
b,s,y,t,l,f,c,b,g,e,c,s,s,w,w,p,w,o,p,n,n,m  
x,f,y,t,l,f,w,n,w,t,b,s,s,w,w,p,w,o,p,n,v,d  
b,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,s,g  
x,y,w,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,n,g  
x,y,n,t,a,f,c,b,w,e,r,s,y,w,w,p,w,o,p,k,s,g  
b,y,y,t,a,f,c,b,w,e,c,s,s,w,w,p,w,o,p,k,s,m  
s,f,g,f,n,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,v,u  
f,s,n,f,n,f,w,b,k,t,e,s,s,w,w,p,w,o,e,k,a,g  
f,f,y,t,l,f,w,n,p,t,b,s,s,w,w,p,w,o,p,n,v,d  
f.s.v.t.l.f.w.n.n.t.h.s.s.w.w.n.w.n.n.v.d

# Processing data with Matlab

● Loading data in Matlab :

● `>> d=dataset('File','data.csv','Delimiter','');`

● `>> [rr,cc]=size(d)`

rr =

4515

cc =

23

# Processing data with Matlab

● Retrieve the class of the 5<sup>th</sup> instance:

● `>> d{5,1}`

ans =

e

● To access the 4<sup>th</sup> attribute of the 7<sup>th</sup> instance:

● `>> d{7,5}`

ans =

x

# Processing data with Matlab

• What are the possible values for 3<sup>rd</sup> attribute ?

• `>> f=unique(d(:,4));`

• `>> size(f,1)`

`ans =`

`8`

`>> f{1,1}`

`ans =`

`b`

`>> f{2,1}`

`ans =`

`c`



# Processing data with Matlab

• What's the probability of **b** across the 3<sup>rd</sup> attribute ?

• `>> [rr,cc]=size(d)`

```
rr =  
    4514
```

```
cc =  
    23
```

• `>> inc=0;`

• `>> for i=1:size(d,1)`

```
    if d{i,4}=='b'  
        inc=inc+1;
```

```
    end
```

```
end
```

• `>> inc`

```
inc =  
    95
```

• `>> prob=inc/size(d,1)`

```
prob =  
    0.0210
```

# Processing data with Matlab

● Using Hamming Distance, how close instance 8 and 18 ?

● `>> sim=0;`

● `>> for i=2:size(d,2)`  
    `if d{8,i}~=d{18,i}`  
    `sim=sim+1;`  
    `end`  
`end`

● `>> sim`  
    `sim =`  
    `5`

# Processing data with Matlab

- Convert categorical data to numerical values:

```
>> d=dataset('File','data.csv','Delimiter','');
```

- Find unique values across a certain attribute:

```
>> h=unique(d(:,2));
```

- Writing a function to convert categorical values to numbers

```
function res=convert2num(u, t)
clear convert2num;
for i=1:size(u,1)
    if u{i,1}==t
        res=i;
        return;
    end
end
end
```

# Processing data with Matlab

• Write a function to convert a column:

```
function res=convertdata2num(at);  
clear convertdata2num;  
h=unique(at);  
for i=1:size(at,1)  
    at{i,1}=convert2num(h,at{i,1});  
end  
res=at;
```

Convert the data now:

```
for i=2:size(d,2)  
    tmp=d(:,i);  
    tmp=convertdata2num(tmp);  
    d(:,i)=tmp;  
end
```

# Classification for Testing Dataset

x,s,n,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,k,s,u → p  
x,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,k,n,g  
x,y,y,t,l,f,c,b,g,e,c,s,s,w,w,p,w,o,p,n,n,g  
s,f,g,f,n,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,y,u  
b,s,y,t,a,f,c,b,k,e,c,s,s,w,w,p,w,o,p,n,s,m  
f,s,w,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,n,v,g  
b,s,y,t,l,f,c,b,g,e,c,s,s,w,w,p,w,o,p,n,n,m  
x,f,y,t,l,f,w,n,w,t,b,s,s,w,w,p,w,o,p,n,v,d  
b,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,s,g  
x,y,w,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,n,g  
x,y,n,t,a,f,c,b,w,e,r,s,y,w,w,p,w,o,p,k,s,g  
b,y,y,t,a,f,c,b,w,e,c,s,s,w,w,p,w,o,p,k,s,m  
s,f,g,f,n,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,v,u  
f,s,n,f,n,f,w,b,k,t,e,s,s,w,w,p,w,o,e,k,a,g  
f,f,y,t,l,f,w,n,p,t,b,s,s,w,w,p,w,o,p,n,v,d  
f.s.v.t.l.f.w.n.n.t.h.s.s.w.w.n.w.o.n.n.v.d

→ e

# Going through the test instances

- Let's classify the instances alternatively between **e** and **p**

```
>> test=dataset('File','test.csv','Delimiter',',');
```

```
>> for i=1:size(test,1)
    if mod(i,2)==0
        res=[res;'e'];
    else
        res=[res;'p'];
    end
end
```

```
>> fid=fopen('myfile2.txt','w');
```

```
>> fprintf(fid,'%c\n',res);
```

```
>> fclose(fid);
```

# Writing your results to a text file

```
• >> g=['e']
```

```
g =
```

```
e
```

```
• >> g=[g;'e']
```

```
g =
```

```
e
```

```
e
```

```
• >> g=[g;'p']
```

```
g =
```

```
e
```

```
e
```

```
p
```

```
• >> myf=fopen('myresult.txt','w');
```

```
• >> fprintf(myf, '%c\n', g);
```

```
• >> fclose(myf);
```

# Submitting your results online

- Visit the following address :
  - [www.imed.ws/machinelearning1](http://www.imed.ws/machinelearning1)
- Type your first name, last name and Email address:
- Choose your **result.txt** file and click Upload
- The score would be shown to you instantly.



# Deadline

● **Deadline is on :**

**Wednesday 11 November 2015**

● **What's further needed :**

● **Report written to explain what you did + copy and paste your code into the report document**