



# Differential Diagnosis of Insomnia/Sleep Apnea Using Sleep Questionnaires

## Workshop

Albert C. Yang, M.D., Ph.D.

Institutes of Brain Science/Digital Medicine Center  
National Yang-Ming University

Apr 23, 2020

[accyang@gmail.com](mailto:accyang@gmail.com)

# PSQI Dataset

- Sleep Center at Taipei Veterans General Hospital
- 200 deidentified, PSG-verified patients with primary insomnia (N=100) and obstructive sleep apnea (N=100)
- Pittsburgh Sleep Quality Index

**Strictly use within the class!**

# PSQI Dataset

- Sleep Center at Taipei Veterans General Hospital
- 200 deidentified, PSG-verified patients with primary insomnia (N=100) and obstructive sleep apnea (N=100)
- Pittsburgh Sleep Quality Index

**Strictly use within the class!**

# PSQI Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	SG Diagnos	PSQI_5.1	PSQI_5.2	PSQI_5.3	PSQI_5.4	PSQI_5.5	PSQI_5.6	PSQI_5.7	PSQI_5.8	PSQI_5.9	PSQI_5.10	sleep qual	sleep latency	sleep durat	sleep efficien	disturbance	medicatio	me dysfun	PSQI_sum
2	PI	2	3	0	2	0	0	0	2	0	0	2	2	3	3	1	0	2	13
3	PI	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	2
4	PI	2	2	0	2	2	0	2	0	0	0	2	2	3	2	1	0	1	11
5	PI	3	3	2	0	0	0	3	3	0	2	2	3	3	0	2	2	1	13
6	PI	3	2	1	0	0	0	0	1	0	0	2	3	0	0	1	3	1	10
7	PI	3	2	2	2	3	3	3	3	3	3	3	3	2	3	3	3	3	20
8	PI	0	3	3	3	2	2	2	0	2	0	2	0	0	0	3	0	2	5
9	PI	2	2	2	1	1	2	2	3	0	0	2	3	0	0	2	0	1	8
10	PI	0	3	2	0	0	0	3	2	0	0	2	0	3	2	2	3	3	15
11	PI	2	2	2	0	0	1	2	3	0	0	1	3	0	0	2	3	1	10
12	PI	3	2	0	0	0	2	2	2	1	0	2	3	0	2	1	0	0	8
13	PI	3	2	2	0	0	0	2	2	0	0	1	3	3	0	1	3	1	12
14	PI	1	3	3	0	3	0	1	3	0	0	2	1	1	1	2	2	1	10
15	PI	3	3	3	0	0	0	2	3	2	0	3	3	3	3	2	3	2	19
16	PI	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1
17	PI	3	3	2	0	0	1	2	2	0	0	2	2	1	3	2	0	2	12
18	PI	1	1	1	0	0	0	0	3	0	0	2	1	0	1	1	3	0	8
19	PI	2	3	2	3	0	2	2	0	0	0	2	2	0	0	2	3	3	12
20	PI	1	2	2	0	0	2	0	2	1	0	3	1	1	3	1	0	1	10
21	PI	3	1	0	0	0	0	2	3	2	0	3	3	1	0	1	3	0	11
22	PI	0	3	3	1	2	2	2	2	3	0	3	0	3	0	2	3	0	11
23	PI	3	3	2	2	3	0	0	2	3	0	2	3	3	2	2	3	2	17
24	PI	3	2	2	1	1	1	1	1	2	0	2	3	3	3	2	3	2	18
25	PI	3	2	3	0	0	0	0	0	0	0	2	2	1	0	1	0	1	7
26	PI	1	2	2	0	0	0	2	2	0	0	1	1	0	0	1	3	3	9
27	PI	3	2	3	0	1	2	2	3	1	0	2	3	3	2	2	3	0	15
28	PI	3	3	3	0	0	0	1	1	0	0	3	3	3	3	1	2	1	16
29	PI	1	1	1	1	1	0	0	0	0	0	2	1	0	0	1	0	1	5
30	PI	3	3	3	0	0	0	3	2	0	0	1	2	1	0	2	0	1	7
31	PI	1	2	0	0	0	3	0	0	2	0	1	1	1	0	1	0	2	6
32	PI	3	1	0	0	0	0	0	3	3	0	2	3	1	1	1	3	3	14

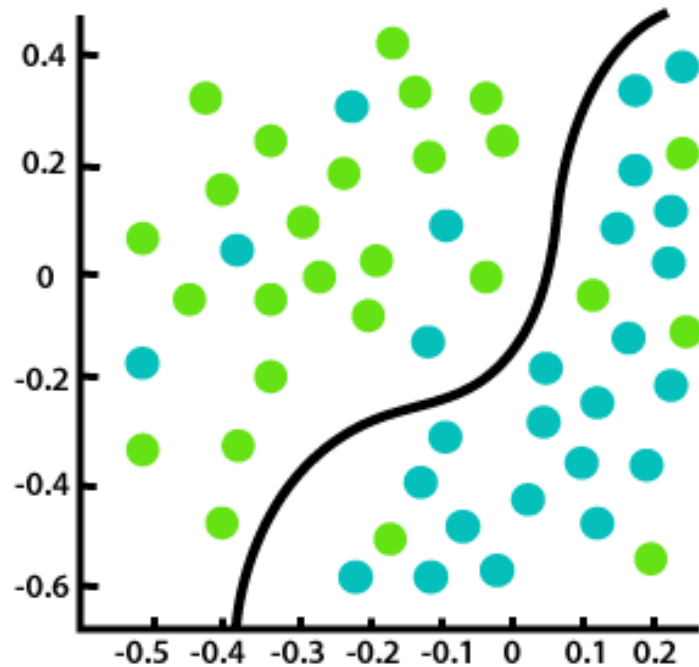
# Import PSQI Data into Matlab

- data =  
readtable('sleep\_questionnaire\_dataset.xlsx');

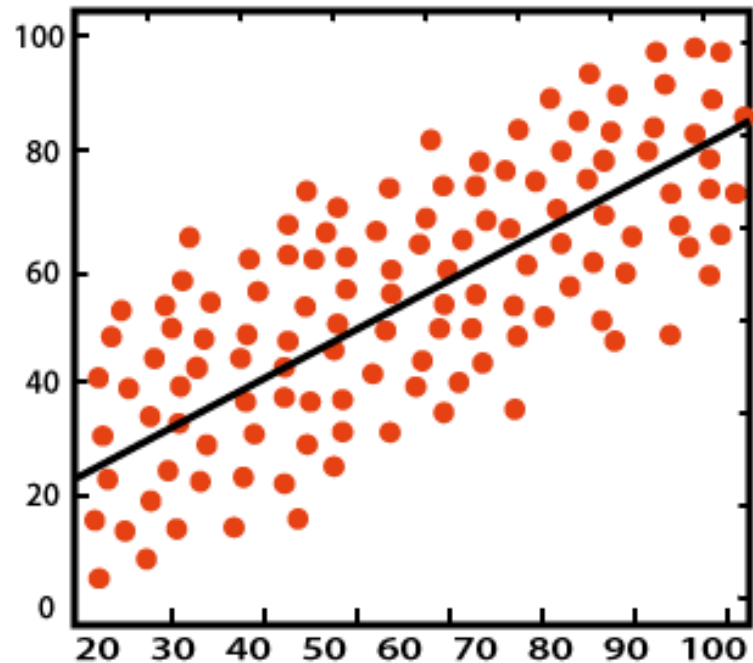
```
>> data = readtable('sleep_questionnaire_dataset.xlsx');
```

Warning: Table variable names were modified to make them valid MATLAB identifiers. The original names are saved in the VariableDescriptions property.

# Classification vs. Regression



Classification

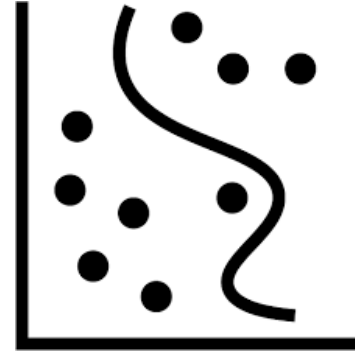


Regression

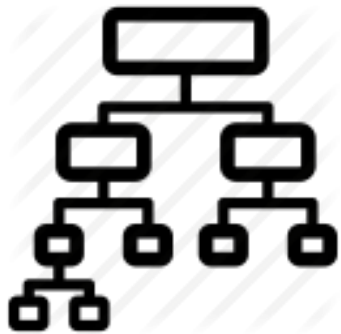
# Classification Methods



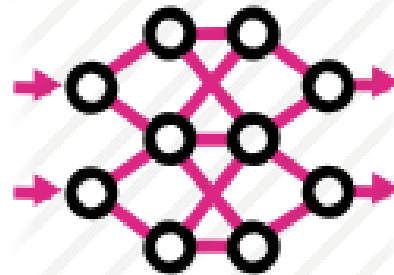
Logistic Regression



Support Vector Machine

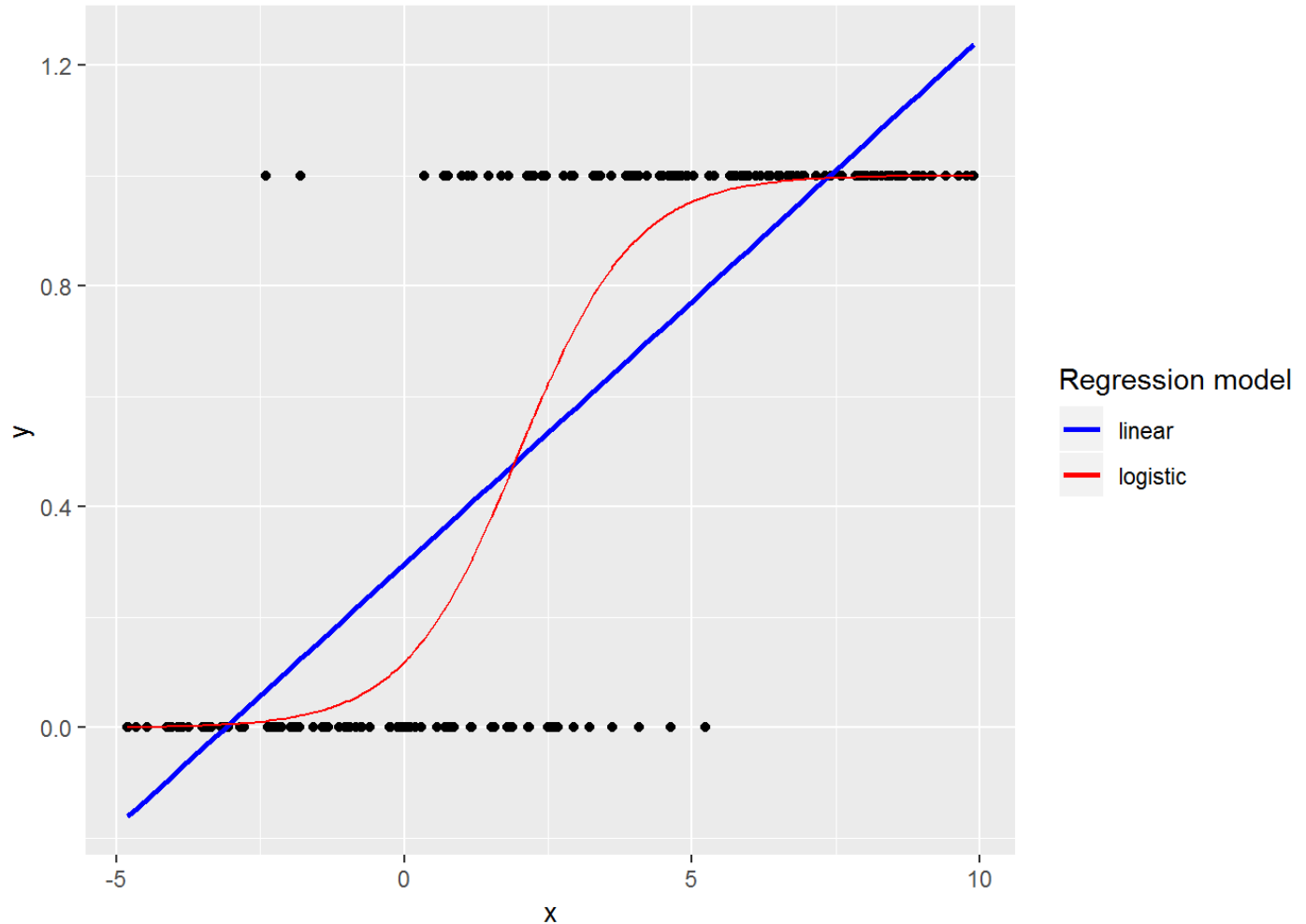


Decision Tree



Neural Network

# Logistic Regression



# Generalized Linear Regression

## Link Function: Logistic

- `predictors = table2array(data(:,2:end));`
- `response = categorical(data.PSGDiagnosis);`
- `model =  
fitglm(predictors,response,'link','logit','Distribution','binomial')`

# Logistic Regression on PSQI Data

## Results

Generalized linear regression model:

$\text{logit}(P(y='SA')) \sim 1 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} +$   
 Distribution = Binomial

Estimated Coefficients:

	Estimate	SE	tStat	pValue
<b>(Intercept)</b>	-0.30246	0.73735	-0.41019	0.68166
<b>κ1</b>	-0.57241	0.37771	-1.5155	0.12965
<b>κ2</b>	-0.27381	0.24661	-1.1103	0.26688
<b>κ3</b>	0.58404	0.24129	2.4205	0.015498
<b>κ4</b>	0.47952	0.23263	2.0613	0.039273
<b>κ5</b>	1.2145	0.22921	5.2985	1.1673e-07
<b>κ6</b>	-0.32949	0.21654	-1.5216	0.12811
<b>κ7</b>	-0.32234	0.22184	-1.453	0.14622
<b>κ8</b>	-0.45843	0.20766	-2.2076	0.027273
<b>κ9</b>	-0.031675	0.2148	-0.14746	0.88277
<b>κ10</b>	0.27626	0.2884	0.9579	0.33811
<b>κ11</b>	-192.82	3.3554e+07	-5.7465e-06	1
<b>κ12</b>	-192.14	3.3554e+07	-5.7263e-06	1
<b>κ13</b>	-192.62	3.3554e+07	-5.7404e-06	1
<b>κ14</b>	-192.24	3.3554e+07	-5.7293e-06	1
<b>κ15</b>	-193.47	3.3554e+07	-5.7659e-06	1
<b>κ16</b>	-192.32	3.3554e+07	-5.7316e-06	1
<b>κ17</b>	-191.84	3.3554e+07	-5.7174e-06	1
<b>κ18</b>	192.51	3.3554e+07	5.7373e-06	1

200 observations, 181 error degrees of freedom

Dispersion: 1

Chi^2-statistic vs. constant model: 92.9, p-value = 4.32e-12

# The Pittsburgh Sleep Quality Index (PSQI)

Instructions: The following questions relate to your usual sleep habits during the past month only. Your answers should indicate the most accurate reply for the majority of days and nights in the past month. Please answer all questions. During the past month,

1. When have you usually gone to bed? \_\_\_\_\_
2. How long (in minutes) has it taken you to fall asleep each night? \_\_\_\_\_
3. When have you usually gotten up in the morning? \_\_\_\_\_
4. How many hours of actual sleep do you get at night? (This may be different than the number of hours you spend in bed) \_\_\_\_\_

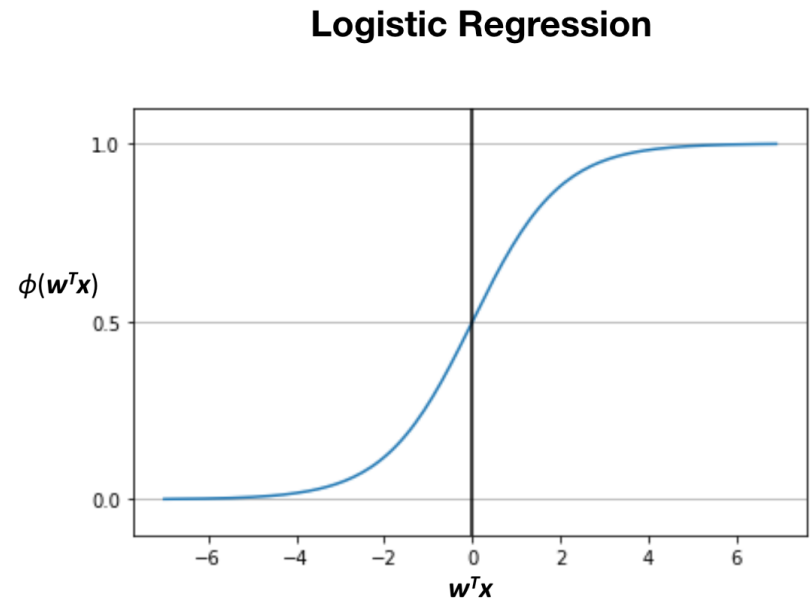
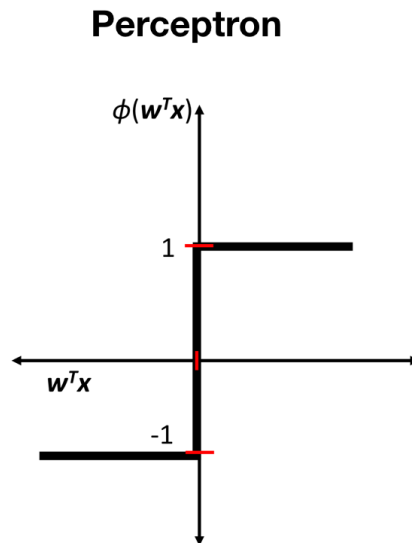
5. During the past month, how often have you had trouble sleeping because you...	Not during the past month (0)	Less than once a week (1)	Once or twice a week (2)	Three or more times a week (3)
a. Cannot get to sleep within 30 minutes				
b. Wake up in the middle of the night or early morning				
c. Have to get up to use the bathroom				
d. Cannot breathe comfortably				
e. Cough or snore loudly				
f. Feel too cold				
g. Feel too hot				
h. Have bad dreams				
i. Have pain				
j. Other reason(s), please describe, including how often you have had trouble sleeping because of this reason(s):				
6. During the past month, how often have you taken medicine (prescribed or "over the counter") to help you sleep?				
7. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?				
8. During the past month, how much of a problem has it been for you to keep up enthusiasm to get things done?				
	Very good (0)	Fairly good (1)	Fairly bad (2)	Very bad (3)
9. During the past month, how would you rate your sleep quality overall?				

# Logistic Regression on PSQI Data

## Evaluate Trained Model

- `ypred = predict(model, predictors);`

ypred	
200x1 double	
	1
1	0.2225
2	0.1157
3	0.6877
4	0.0151
5	0.1562
6	0.7411
7	2.2204e-16
8	0.0701
9	0.1678
10	0.0442
11	0.0085
12	0.0988
13	0.7370
14	0.0734
15	0.6302
16	0.0342
17	0.0895
18	0.4477
19	0.1170
20	0.0090

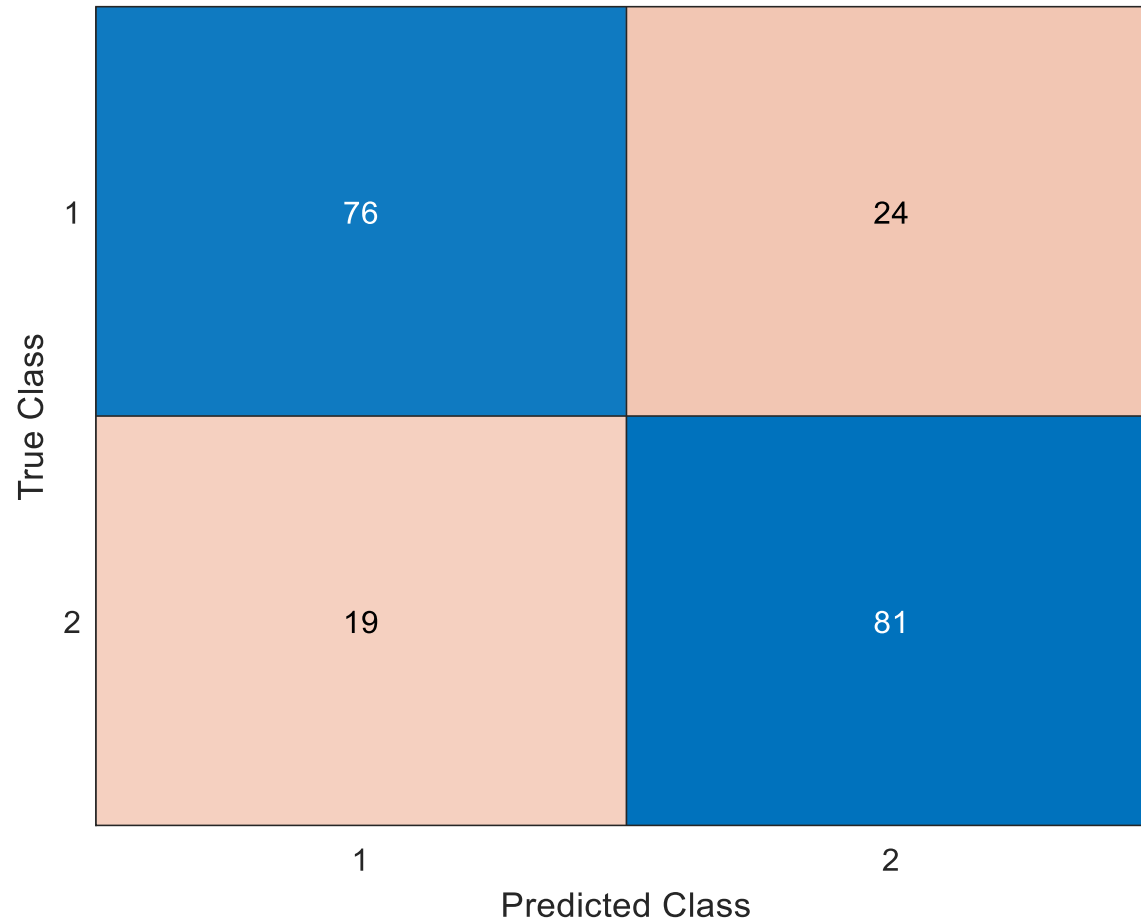


# Logistic Regression on PSQI Data

## Confusion Matrix

- `ypred = ypred >= 0.5;`
- `ypred = double(ypred);`
- `[group_cat, ~, group] = unique(response);`
- `group = group - 1;`
  
- `c = confusionmat(group,ypred);`
- `confusionchart(c)`

# Confusion Matrix



# Confusion Matrix Statistics

		Predicted Class	
		No	Yes
Observed Class	No	TN	FP
	Yes	FN	TP

TN True Negative  
FP False Positive  
FN False Negative  
TP True Positive

## Model Performance

Accuracy =  $(TN+TP)/(TN+FP+FN+TP)$

Precision =  $TP/(FP+TP)$

Sensitivity =  $TP/(TP+FN)$

Specificity =  $TN/(TN+FP)$

Binary Confusion Matrix

# Confusion Matrix Statistics

- `stats = confusionmatStats(group,ypred)`

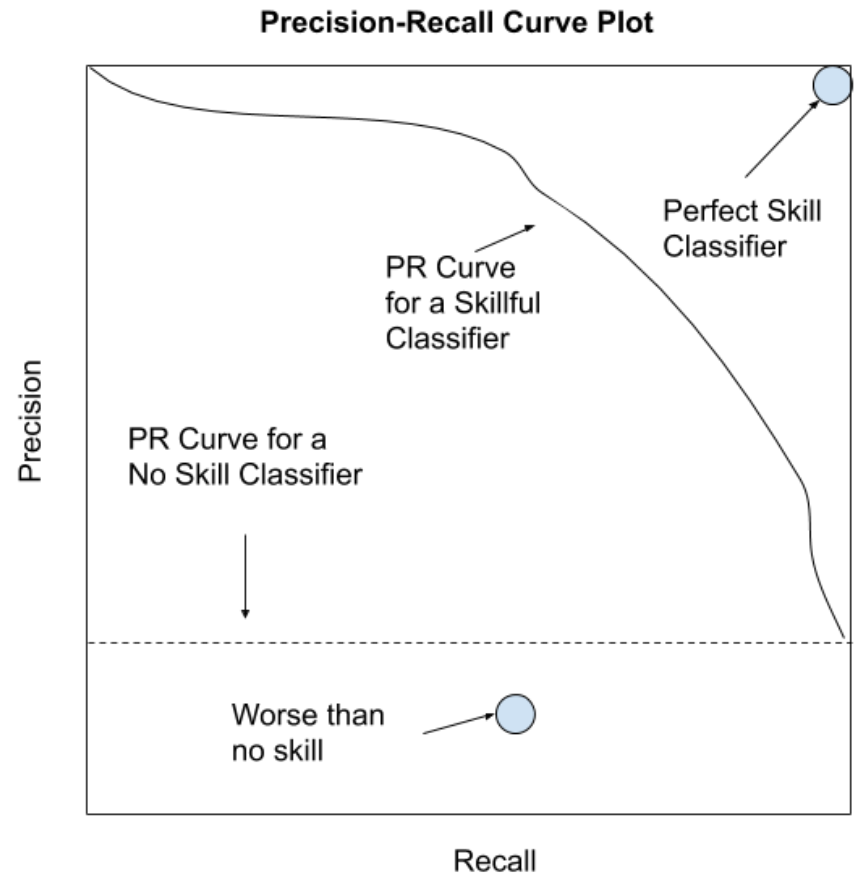
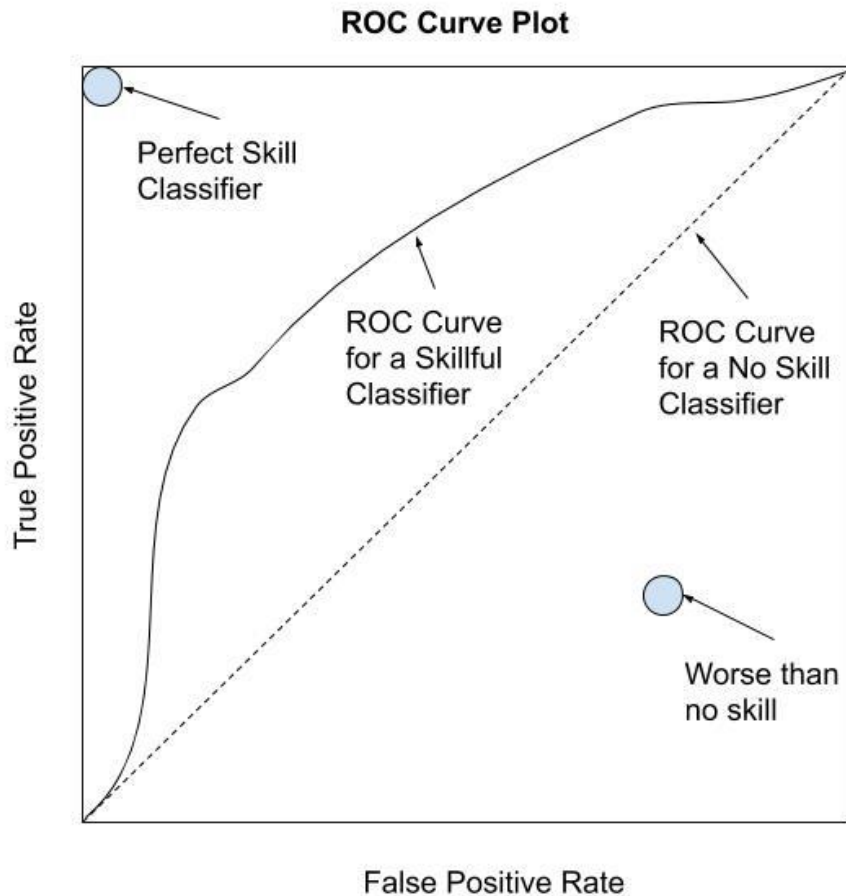
```
>> confusionmatStats(group,ypred)
```

```
ans =
```

```
struct with fields:
```

```
confusionMat: [2x2 double]  
  accuracy: [2x1 double]  
sensitivity: [2x1 double]  
specificity: [2x1 double]  
  precision: [2x1 double]  
   recall: [2x1 double]  
   Fscore: [2x1 double]  
groupOrder: [2x1 double]
```

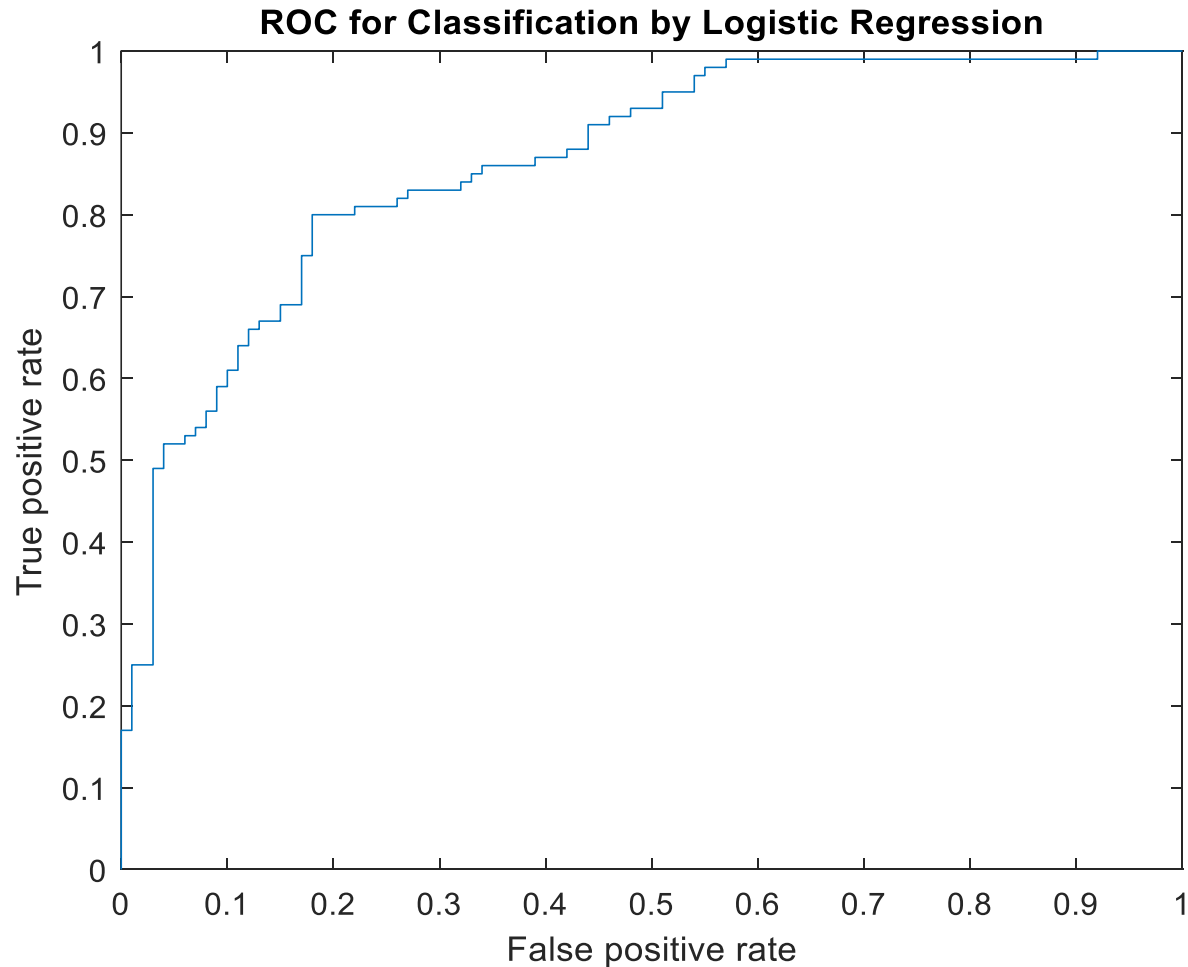
# Receiver Operating Characteristic



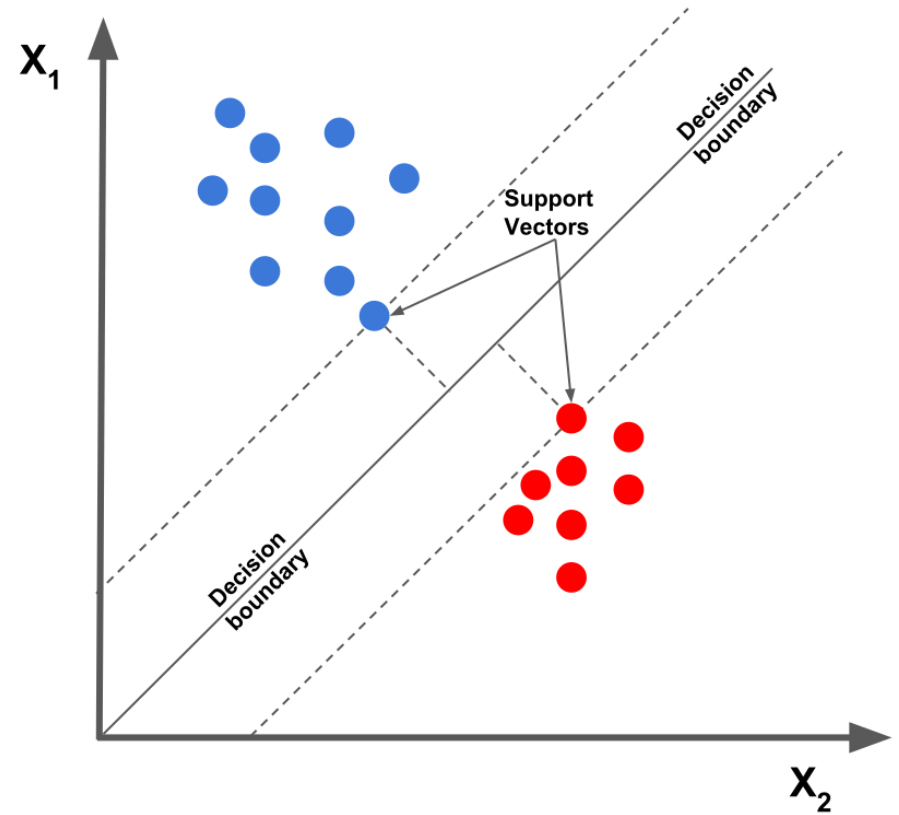
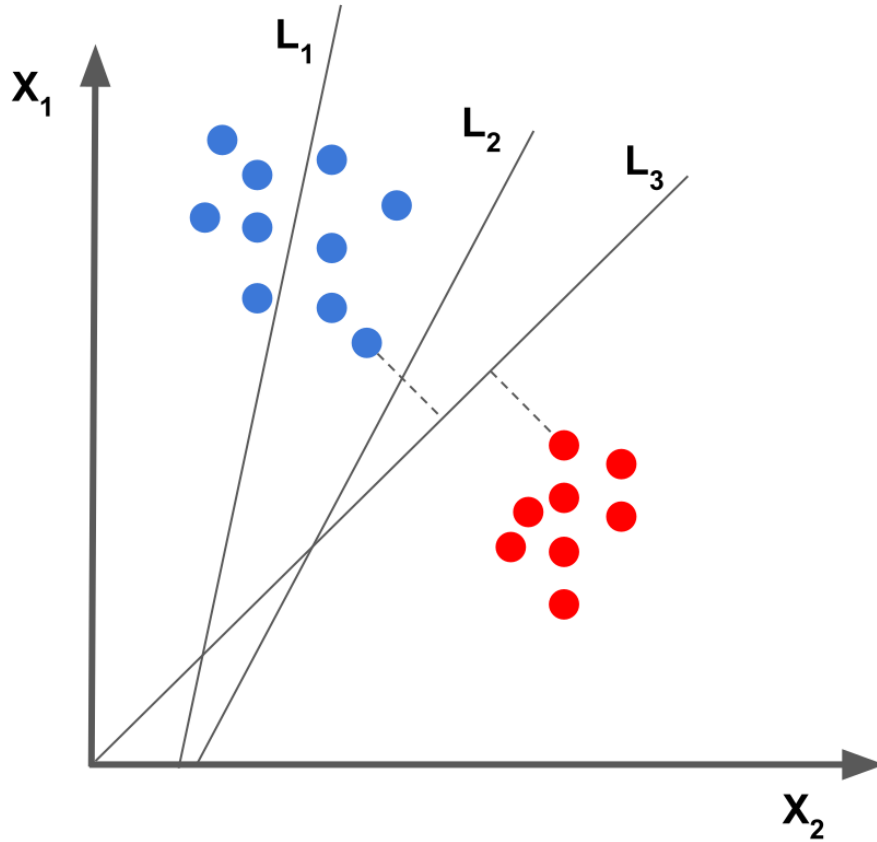
# Receiver Operating Characteristic

- `scores = model.Fitted.Probability;`
- `[X,Y,T,AUC] = perfcurve(group,scores,1);`
- `plot(X,Y)`
- `xlabel('False positive rate')`
- `ylabel('True positive rate')`
- `title('ROC for Classification by Logistic Regression')`

# Receiver Operating Characteristic



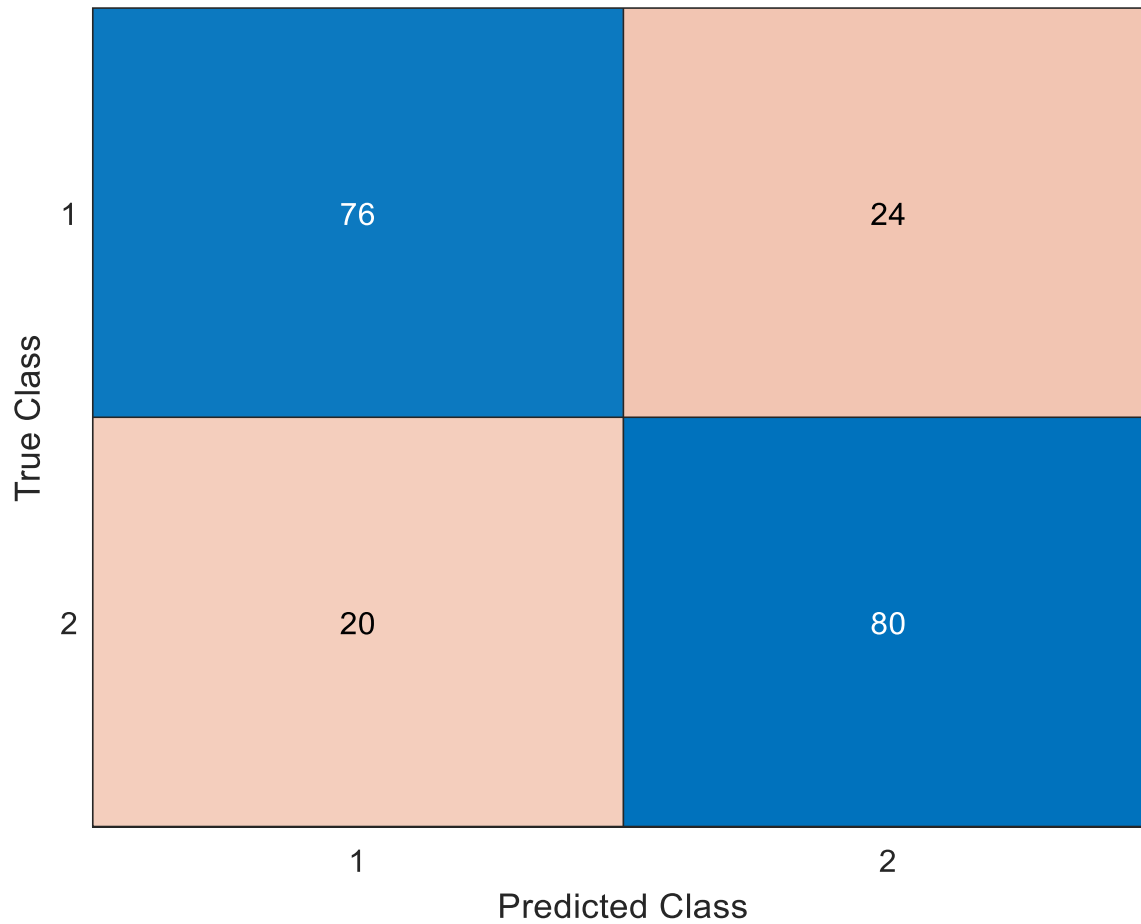
# Support Vector Machine



# Support Vector Machine

- `model1 = fitcsvm(data,'PSGDiagnosis');`
- `ypred1 = predict(model1,predictors);`
- `[group_cat, ~, group_pred] = unique(ypred1);`
- `group_pred = group_pred - 1;`
- `c1 = confusionmat(group,group_pred);`
- `confusionchart(c1)`
- `stats1 = confusionmatStats(group,group_pred)`

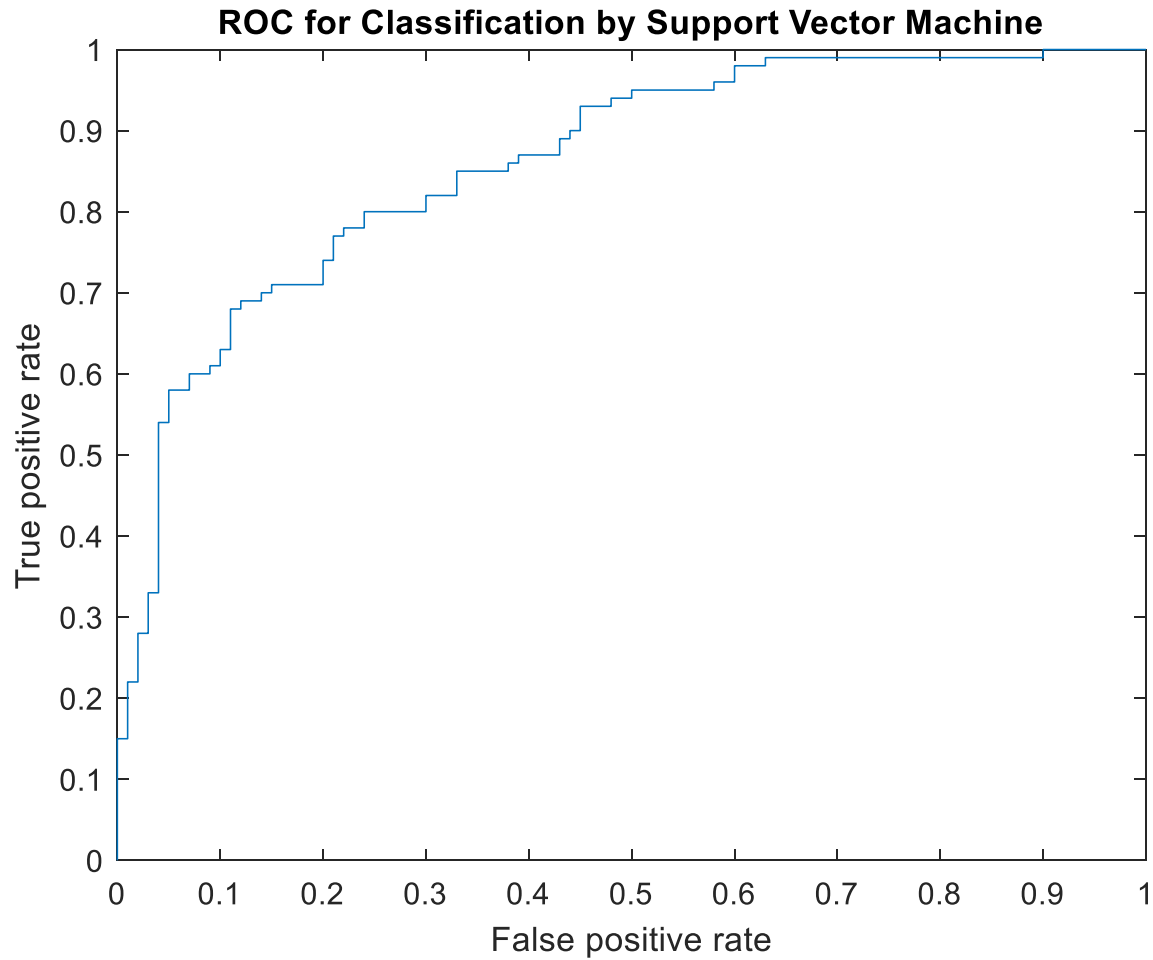
# Confusion Matrix



# Receiver Operating Characteristic

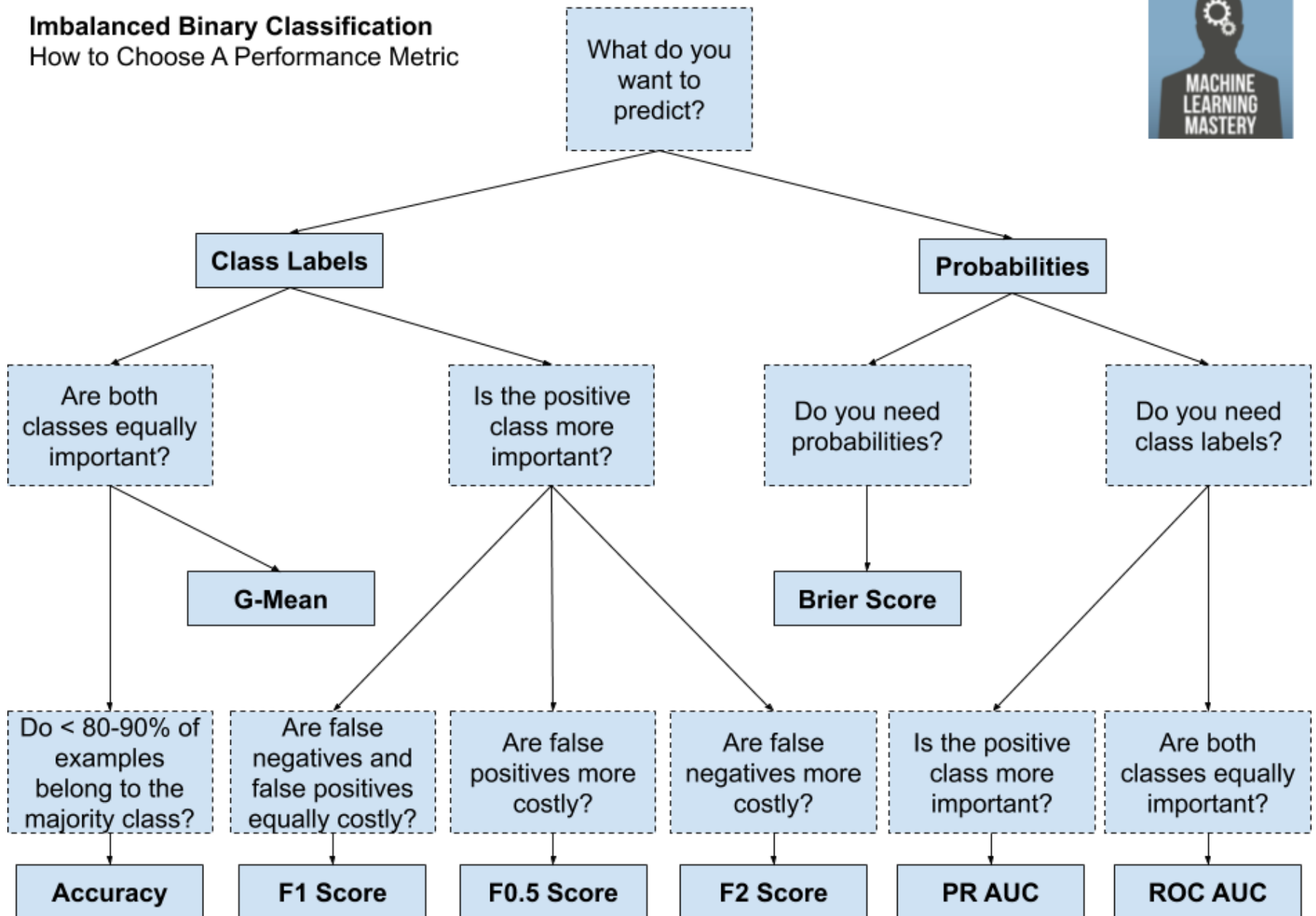
- `model1 = fitPosterior(model1);`
- `[~,score_svm] = resubPredict(model1);`
- `[X,Y,T,AUC] = perfcurve(group,score_svm(:,2),1);`
- `plot(X,Y)`
- `xlabel('False positive rate')`
- `ylabel('True positive rate')`
- `title('ROC for Classification by Support Vector Machine')`

# Receiver Operating Characteristic



# Imbalanced Binary Classification

## How to Choose A Performance Metric



# Online References

- [https://vas3k.com/blog/machine\\_learning/?fbclid=IwAR0NjjOJIZt4-KiaBGi11DskcBHAAa2d6xaUchkPZdDch7pxS5sbcrZkUBJA](https://vas3k.com/blog/machine_learning/?fbclid=IwAR0NjjOJIZt4-KiaBGi11DskcBHAAa2d6xaUchkPZdDch7pxS5sbcrZkUBJA)
- <https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/>
- <https://builtin.com/data-science/supervised-machine-learning-classification>