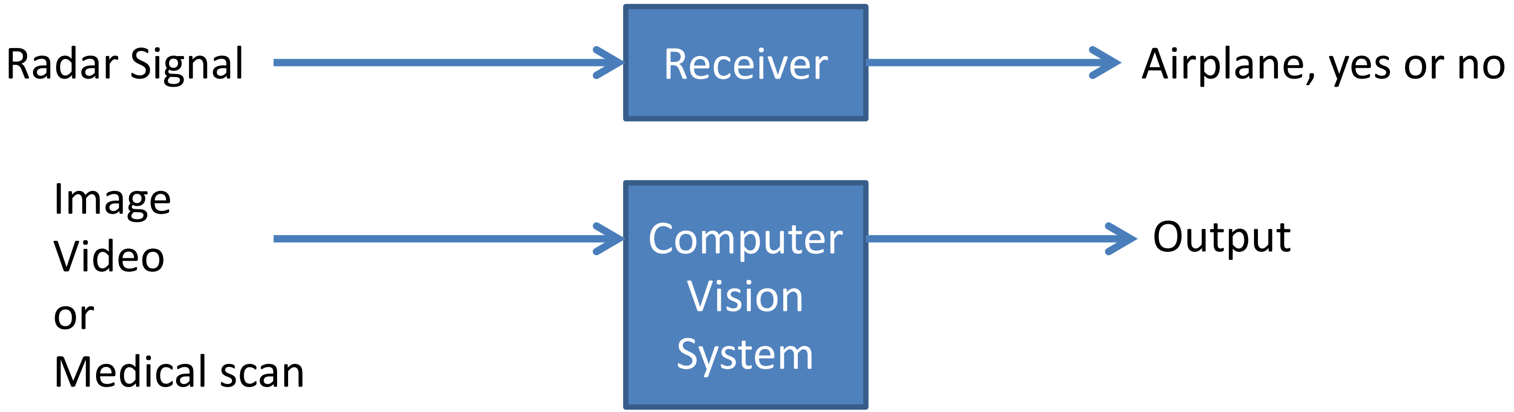
Performance Analysis of Computer Vision Systems

CS 585
Margrit Betke

ROC Analysis

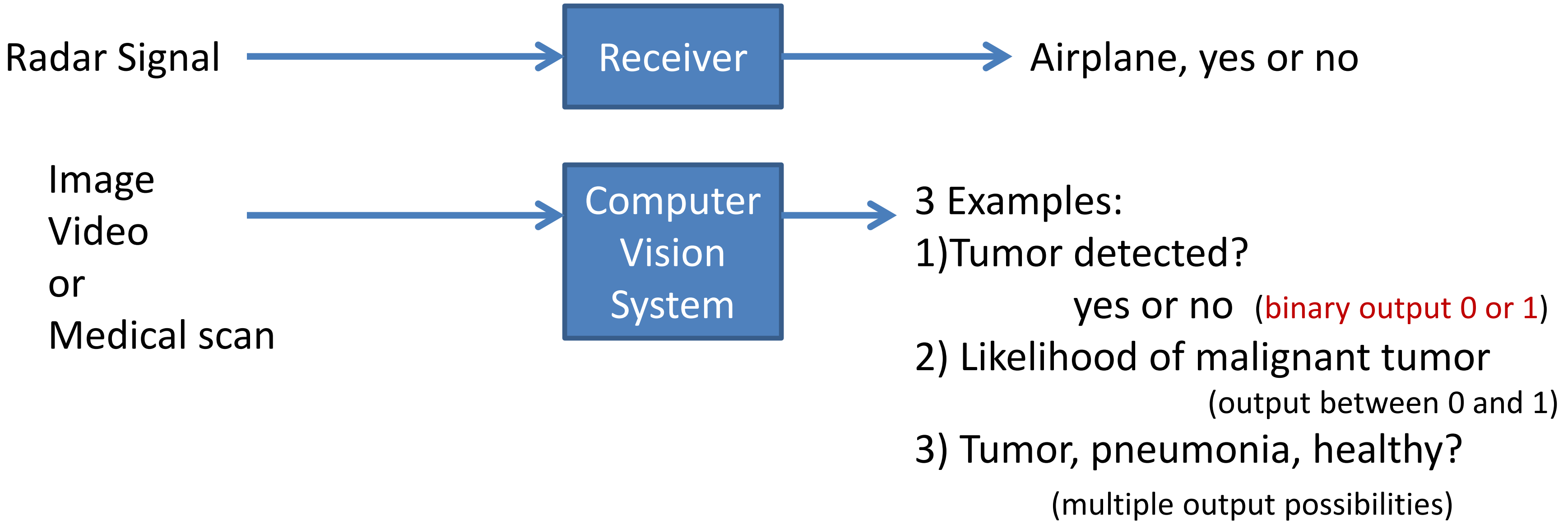
ROC = receiver operating characteristics (historic name from radar signal processing)
ROC Analysis = Method to organize, visualize, and evaluate results of an computer vision system



ROC Analysis

ROC = receiver operating characteristics (historic name from radar signal processing)

ROC Analysis = Method to organize, visualize, and evaluate results of an computer vision system



ROC Analysis

ROC = receiver operating characteristics   (historic name from radar signal processing)
ROC Analysis = Method to organize, visualize, and evaluate results of an computer vision system

Radar Signal → Receiver → Airplane, yes or no

Image
Video
or
Medical scan
→ Computer Vision System →

3 Examples:

1)Tumor detected?
yes or no (binary output 0 or 1)

2) Likelihood of malignant tumor
(output between 0 and 1)

3) Tumor, pneumonia, healthy?
(multiple output possibilities)

Binary Classifier

Multi-class Classifier

Confusion Matrix for Binary Classifier Case

"Truth" =
Ground truth =
Gold standard =
Actual class

Vision System output =
Hypothesis =
Predicted class

	1	0
1	True Positive (TP)	False Positive (FP)
0	False Negative (FN)	True Negative (TN)

Confusion Matrix for Binary Classifier Case

Example with 20 samples

"Truth" =
Ground truth =
Gold standard =
Actual class

AI System output =
Hypothesis =
Predicted class

	1	0
1	True Positive (TP): 6	False Positive (FP): 4
0	False Negative (FN): 2	True Negative (TN): 8

Confusion Matrix for Binary Classifier Case

"Truth" =
Ground truth =
Gold standard =
Actual class

AI System output =
Hypothesis =
Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

1st step of analyzing the confusion matrix:
Check that sum of matrix entries = number of samples used to test AI system

Confusion Matrix for Binary Classifier Case

"Truth" = Ground truth = Gold standard = Actual class

System output = Hypothesis = Predicted class

	1	0
1	True Positive (TP): 6	False Positive (FP): 4
0	False Negative (FN): 2	True Negative (TN): 8

Good System?

Confusion Matrix for Binary Output Case

"Truth" =
Ground truth =
Gold standard =
Actual class

System output =
Hypothesis =
Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Good System? We want high values in diagonal of matrix.

TP+TN=6+8=14

Confusion Matrix for Binary Output Case

TP+TN=6+8=14

Total number of samples = 20

"Truth" =
Ground truth =
Gold standard =
Actual class

System output =
Hypothesis =
Predicted class

	1	0
1	True Positive (TP): 6	False Positive (FP): 4
0	False Negative (FN): 2	True Negative (TN): 8

2nd step of analyzing the confusion matrix:
Compute sum of diagonal entries and compare that with total number of samples

Confusion Matrix for Binary Output Case

TP+TN=6+8=14

Total number of samples = 20

14 versus 20: Is this a good system?

"Truth" =
Ground truth =
Gold standard =
Actual class

System output =
Hypothesis =
Predicted class

	1	0
1	True Positive (TP): 6	False Positive (FP): 4
0	False Negative (FN): 2	True Negative (TN): 8

2^{nd} step of analyzing the confusion matrix:

Compute sum of diagonal entries and compare that with total number of samples

Confusion Matrix for Binary Output Case

TP+TN=6+8=14

Total number of samples = 20

Accuracy of System:

14/20 = 0.7

"Truth" = Ground truth = Gold standard = Actual class

System output = Hypothesis = Predicted class

	1	0
1	True Positive (TP): 6	False Positive (FP): 4
0	False Negative (FN): 2	True Negative (TN): 8

2nd step of analyzing the confusion matrix:

Compute sum of diagonal entries and compare that with total number of samples

Confusion Matrix for Binary Output Case

"Truth" =
Ground truth =
Gold standard =
Actual class

System output =
Hypothesis =
Predicted class

	1	0
1	True Positive (TP): 6	False Positive (FP): 4
0	False Negative (FN): 2	True Negative (TN): 8

Positive samples =
TP+FN =
8

Negative samples =
FP+FN =
12

How sensitive is the classifier in finding the positives?

true positive rate = **tp** =
TP/(TP+FN) = 6/8 = ¾
= ***recall*** = ***sensitivity***

e.g.
How sensitive to tumor?

"Truth" =
Ground truth =
Gold standard =
Actual class

System output =
Hypothesis =
Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Positive samples =
P= TP+FN =
8

false positive rate = **fp** = **FP/(FP+TN) = 4/12 = 1/3**

"Truth" = Ground truth = Gold standard = Actual class

System output = Hypothesis = Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

e.g., overcalling tumors

Negative samples = FP+TN = 12

true positive rate = **tp** =
TP/(TP+FN) = 6/8 = ¾
= ***recall*** = ***sensitivity***

false positive rate = **fp** =
FP/(FP+TN) = 4/12 = 1/3

"Truth" =
Ground truth =
Gold standard =
Actual class

System output =
Hypothesis =
Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Positive samples =
TP+FN =
8

Negative samples =
FP+TN =
12

How specific is the classifier in finding the negatives?

Instead of *fp*, we sometimes focus on

1 – *fp* = *specificity*
TN/(FP+TN) = 8/12 = 2/3

e.g., model not specific to tumors if lots of FPs compared non-tumor examples

"Truth" = Ground truth = Gold standard = Actual class

System output = Hypothesis = Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Negative samples = FP+TN = 12

true positive rate = tp =
TP/(TP+FN) = 6/8 = ¾
= ***recall*** = ***sensitivity***

precision =
TP/(TP+FP) = 6/10 = 3/5

e.g., how precise is the classifier in finding the tumors?

"Truth" =
Ground truth =
Gold standard =
Actual class

System output =
Hypothesis =
Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Positive hypotheses = **TP**+**FP** = 10

Positive samples = TP+FN = 8

F1 Score

true positive rate **= tp =**
TP/(TP+FN) = 6/8 = ¾
= ***recall*** **=** ***sensitivity***

F1 score = 2 recall x precision/(recall + precision)
= 2 x 3/4 x 3/5 / (3/4+3/5) = 2/3

precision **=**
TP/(TP+FP) = 6/10 = 3/5

System output =
Hypothesis =
Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Positive hypotheses = **TP**+**FP** = 10

Positive samples = TP+FN = 8

F1 Score

true positive rate **= tp = TP/(TP+FN) = 6/8 = ¾ = 0.75 =** ***recall*** **=** ***sensitivity***

precision **= TP/(TP+FP) = 6/10 = 3/5 = 0.6**

F1 score = 2 recall x precision/(recall + precision)
= 2 x 3/4 x 3/5 / (3/4+3/5) = 2/3 = **0.667**

System output = Hypothesis = Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Balanced Accuracy

true positive rate **= tp =**
TP/(TP+FN) = 6/8 = ¾ = 0.75
= *recall* = *sensitivity*

Balanced Accuracy = (((TP/(TP+FN)+(TN/(TN+FP))) / 2 =
(sensitivity + specificity)/2 =
(3/4 + 2/3)/2 = (0.75 + 0.67)/2 = 2/3 = **0.708**

1 – *fp* = *specificity*
TN/(FP+TN) = 8/12 = 0.67

System output =
Hypothesis =
Predicted class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Accuracy vs. F1 Score vs. Balanced Accuracy

Accuracy = (TP+TN)/everything = **0.700**

F1 Score = 2 recall x precision/(recall + precision) = **0.667**

Balanced Accuracy = (sensitivity + specificity)/2 = **0.708**

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Accuracy vs. F1 Score vs. Balanced Accuracy

Accuracy = (TP+TN)/everything = **0.999**
F1 Score = 2 recall x precision/(recall + precision) = **0.667**
Balanced Accuracy = (sensitivity + specificity)/2 = **0.833**

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8000**

Accuracy vs. F1 Score vs. Balanced Accuracy

Accuracy = (TP+TN)/everything = **0.999**

F1 Score = 2 recall x precision/(recall + precision) = **0.999**

Balanced Accuracy = (sensitivity + specificity)/2 = **0.833**

	1	0
1	**True Positive (TP):6000**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

Terms to remember:

ROC

Classifier

Predictor

Likelihood

Ground truth, gold standard
Hypothesis

Accuracy, Balanced Accuracy, F1 score

False positive rate & False negative rate

Recall & Precision

Sensitivity & Specificity

Building an ROC curve for a Vision System: One classifier at time

TP+TN=6+8=14

Total number of samples = 20

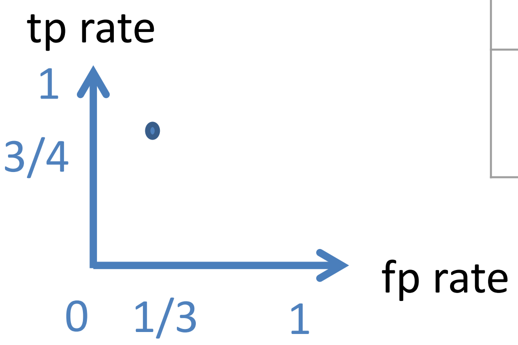
Accuracy of Classifier:

14/20 = 0.7

false positive rate = 1/3

true positive rate = 3/4

ROC curve has 1 point:



"Truth" =
Ground truth =
Gold standard =
Actual class

	1	0
1	**True Positive (TP): 6**	**False Positive (FP): 4**
0	**False Negative (FN): 2**	**True Negative (TN): 8**

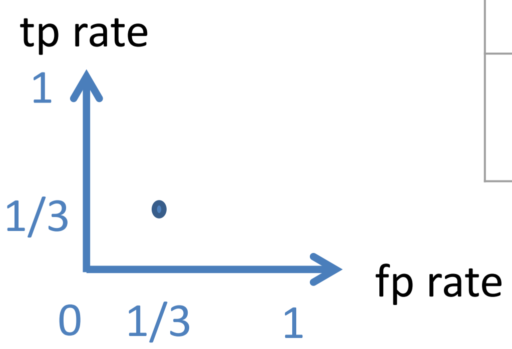
Good Classifier?

"Truth" =
Ground truth =
Gold standard =
Actual class

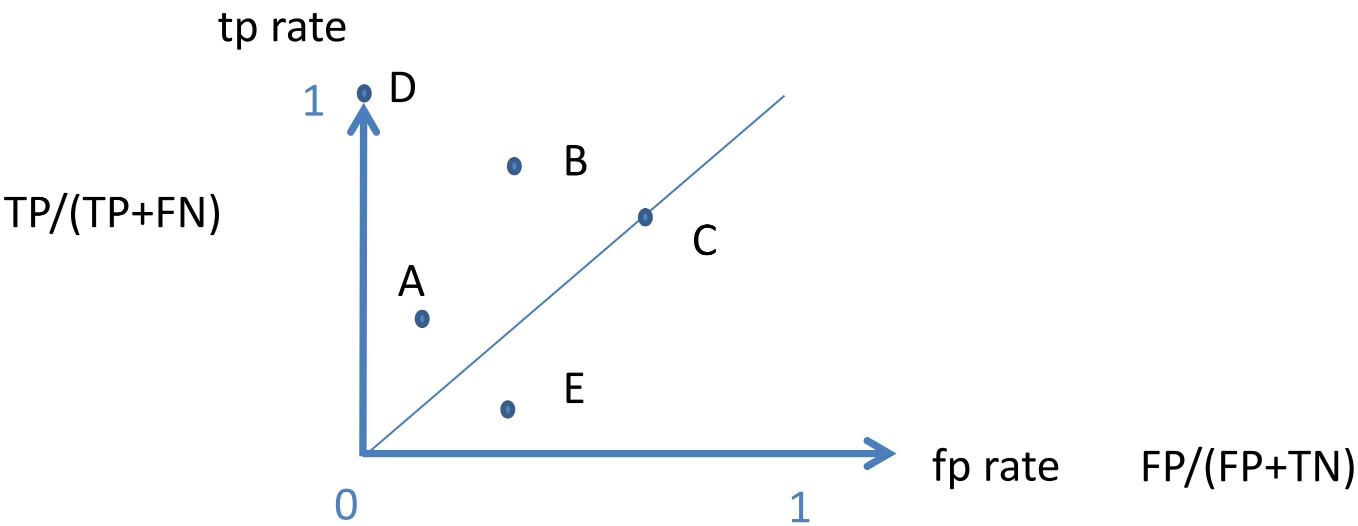
	1	0
1	True Positive (TP): 4	False Positive (FP): 4
0	False Negative (FN): 8	True Negative (TN): 8

false positive rate = 1/3
true positive rate = 1/3

ROC curve has 1 point:



Comparing Classifiers



Classifier A:

Classifier B:

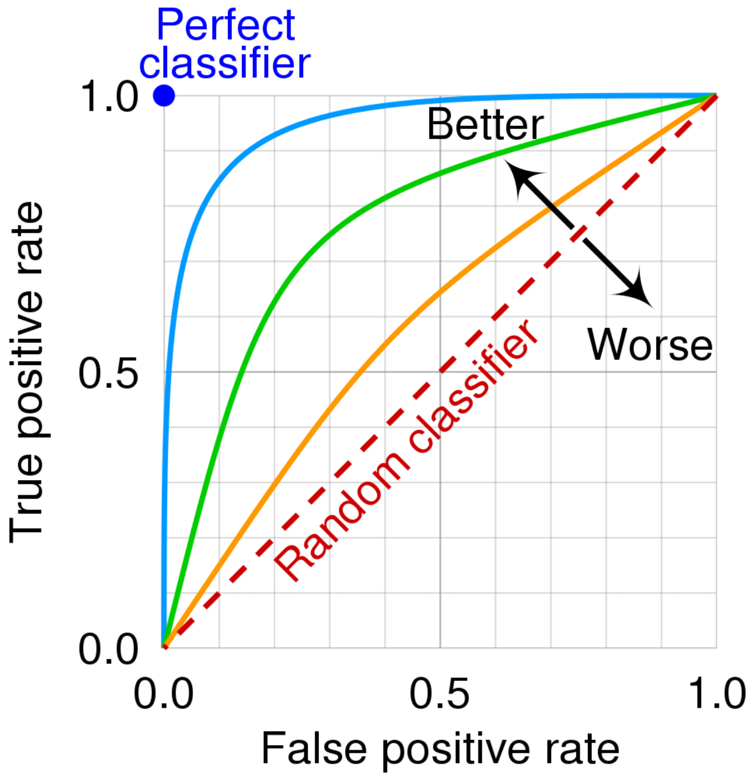
Classifier C:

Classifier D:

Classifier E:

See paper by Fawcett

ROC Curves



Each colored line shows the behavior of a binary classifier when a parameter is changed.

Example for tumor prediction classifier:
The parameter could be the threshold T on circularity of the mass:
circularity > T   Predict "tumor"
<= T   Predict "no tumor"

Plot the curve for by testing performance on T=0.95, 0.9, 0.85, etc

Curves drawn by CMG Lee based onhttps://commons.wikimedia.org/wiki/File:Roc_curve.svg

Confusion Matrix for Multiple Classes

"Truth" = Ground truth = Gold standard = Actual class

System Output = Hypothesis = Predicted class

	Class 1	Class 2	Class 3	Class 4
Class 1	100%	15%	10%	7%
Class 2	0%	80%	10%	3%
Class 3	0%	3%	80%	70%
Class 4	0%	2%	0%	20%

Confusion Matrix for Multiple Classes

"Truth" = Ground truth = Gold standard = Actual class

System Output = Hypothesis = Predicted class

	Class 1	Class 2	Class 3	Class 4
Class 1	100%	15%	10%	7%
Class 2	0%	80%	10%	3%
Class 3	0%	3%	80%	70%
Class 4	0%	2%	0%	20%

Warning about how to talk about Percentage Differences!

Can you say that this classifier's ability to predict class 3 is 20% worse than class 1?

Idea is
100% - 80% = 20%

	Class 1	Class 2	Class 3	Class 4
Class 1	100%	15%	10%	7%
Class 2	0%	80%	10%	3%
Class 3	0%	3%	80%	70%
Class 4	0%	2%	0%	20%

Warning about how to talk about Percentage Differences!

Can you say that this classifier's ability to predict class 1 is 20% better than class 3? NO

Idea is
100% - 80% = 20%

You have to say
20 percent points.

"20 % better" would
add up to 96% because:

80% * 20% = 16%

16%+80% = 96%
not 100% !

	Class 1	Class 2	Class 3	Class 4
Class 1	100%	15%	10%	7%
Class 2	0%	80%	10%	3%
Class 3	0%	3%	80%	70%
Class 4	0%	2%	0%	20%

The difference between percentages is called percent points.

Confusion Matrix for Multiple Classes

Note: Rows and columns of a confusion matrix may be reversed
Reporting only percentages and not actual number is usually NOT a good practice.

Example of a multi-class confusion matrix in one of my papers (Zhang et al, IJCV 2017):

Each row corresponds to a ground-truth category label. The percentage reported is the average proportion of images of the category A (row number) labeled as category B (column number). For over 90% images, predicted labels are consistent with the ground-truth labels.

	0	1	2	3	4+
0	90% (179)	5% (9)	2% (3)	1% (2)	3% (6)
1	1% (2)	96% (191)	3% (5)	1% (1)	1% (1)
2	0	3% (6)	95% (189)	3% (5)	0
3	0	1% (1)	3% (5)	96% (191)	1% (2)
4+	13% (26)	3% (6)	4% (8)	2% (3)	78% (156)

Learning Outcomes

Be able to explain why computer vision scientists do ROC analysis

Explain the concepts classifier, predictor, likelihood, binary classifier, multi-class classifier, ground truth, gold standard, hypothesis, confusion matrix

Be able to define and apply formulas for computing the false positive rate, false negative rate, recall & precision, sensitivity & specificity, accuracy, balanced accuracy, F1 score

Know to analyze the confusion matrix and select appropriate performance measures to evaluate the computer vision system **(including your homework 2 system)**

Be able to draw and analyze a ROC curve for a parameterized classifier problem

Be able to compare ROC curves and points on the curves to select the most appropriate classifier

Remember the different meaning of percentage and percentage point

Remember not to report only percentages for a computer vision system, but the actual numbers

Know the difference between a confusion matrix for a binary classifier and a multi-class classifier