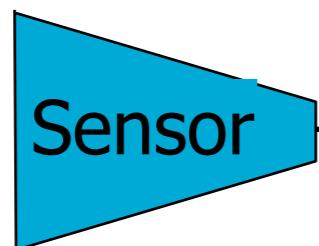
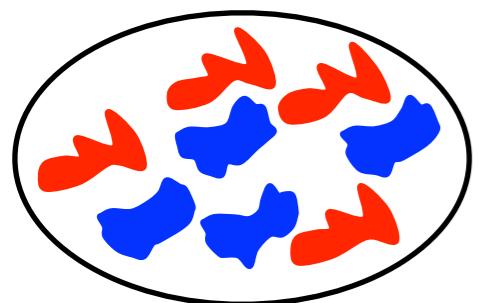


Classifier evaluation

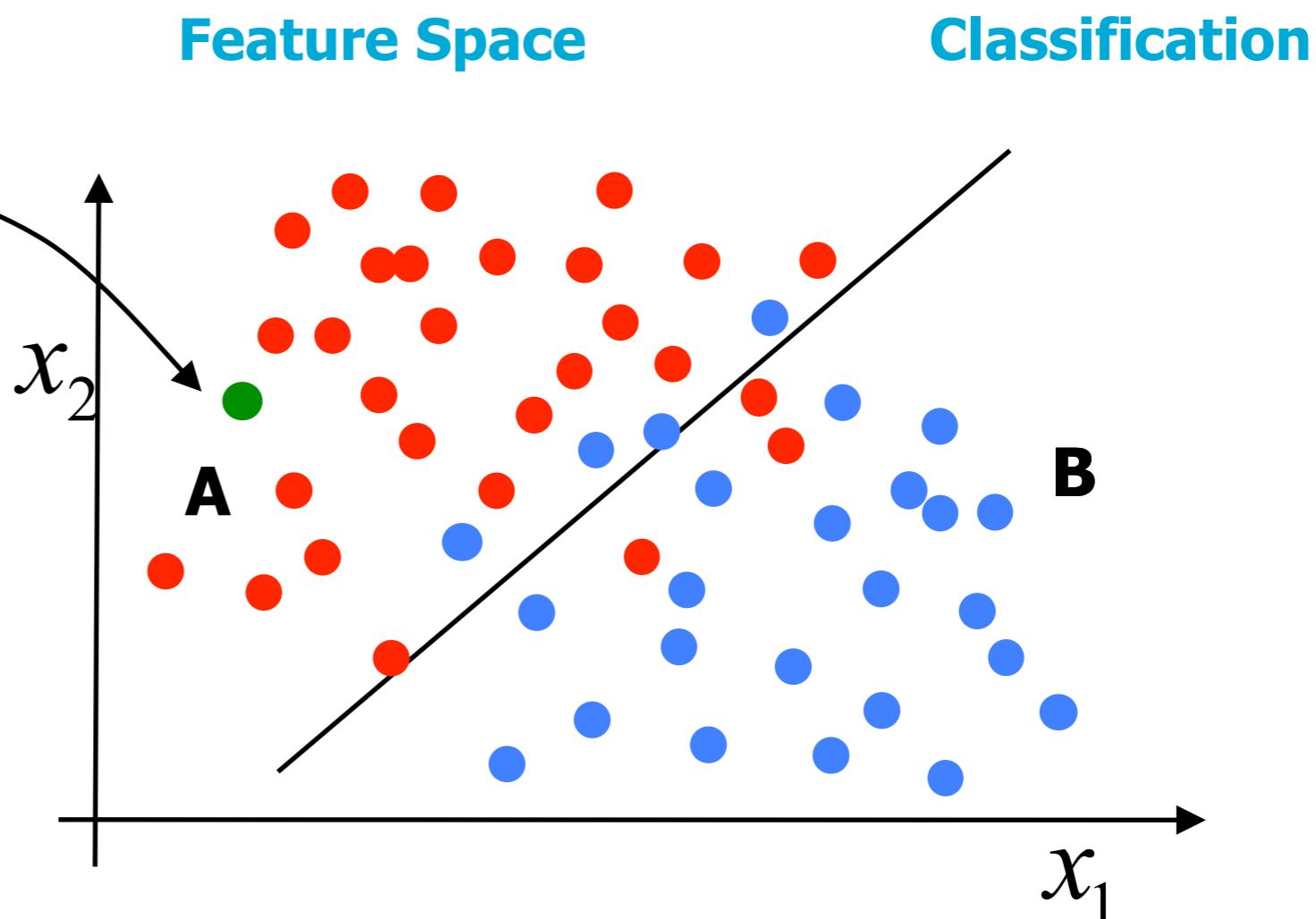


Representation

Generalization

 Test object
classified as 'A'

- Confusion matrices**
- Rejection curves**
- 2-class errors**
- ROC curves**



Classifier evaluation

- This lecture:
 - Train, validation and test set
 - Learning curve
 - Bootstrapping, cross-validation, leave-one-out
 - Classifier complexity
 - Feature curves
- Second lecture:
 - Confusion matrices,
 - Reject curve
 - ROC curve
- Question hour

Confusion Matrices

- Provides counts of class-dependent errors : How many object have been classified as A that should have been classified as B?
- Give a more detailed view than overall error rate
- Can be used to estimate overall cost for particular classifier

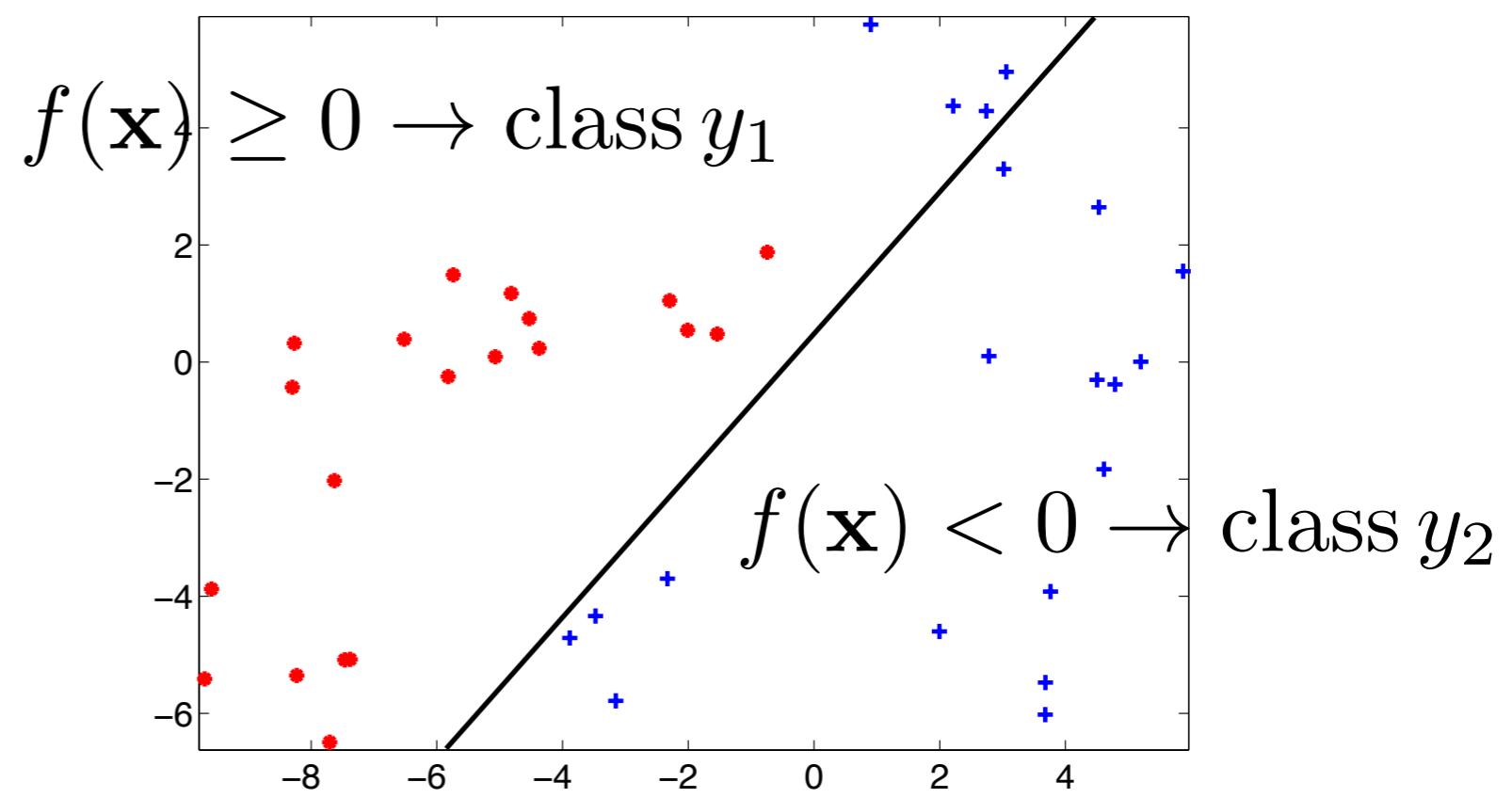
Confusion Matrix (1)

Real
labels:

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \dots \\ \lambda_N \end{bmatrix}$$

Predicted
labels:

$$L = \begin{bmatrix} l_1 \\ \dots \\ l_N \end{bmatrix}$$



Confusion matrix:

$$C = \begin{bmatrix} c_{11} & \dots & c_{1K} \\ \dots & \dots & \dots \\ c_{K1} & \dots & c_{KK} \end{bmatrix}$$

$$c_{ij} = N \cdot P[l_j | \lambda_i]$$

Confusion Matrix (2)

$$N_A = 10, N_B = 30, N_C = 20$$

$$E = \frac{c_{12} + c_{13} + c_{21} + c_{23} + c_{31} + c_{32}}{N_A + N_B + N_C}$$

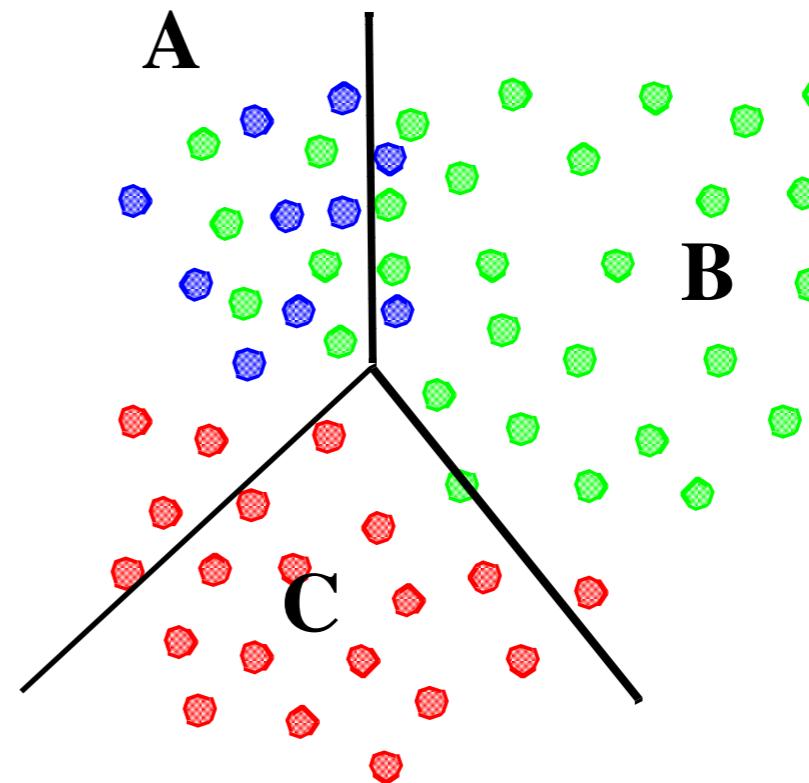
$$E = 14 / 60 = 0.2333$$

$C = \text{confmat}(\Lambda, L)$

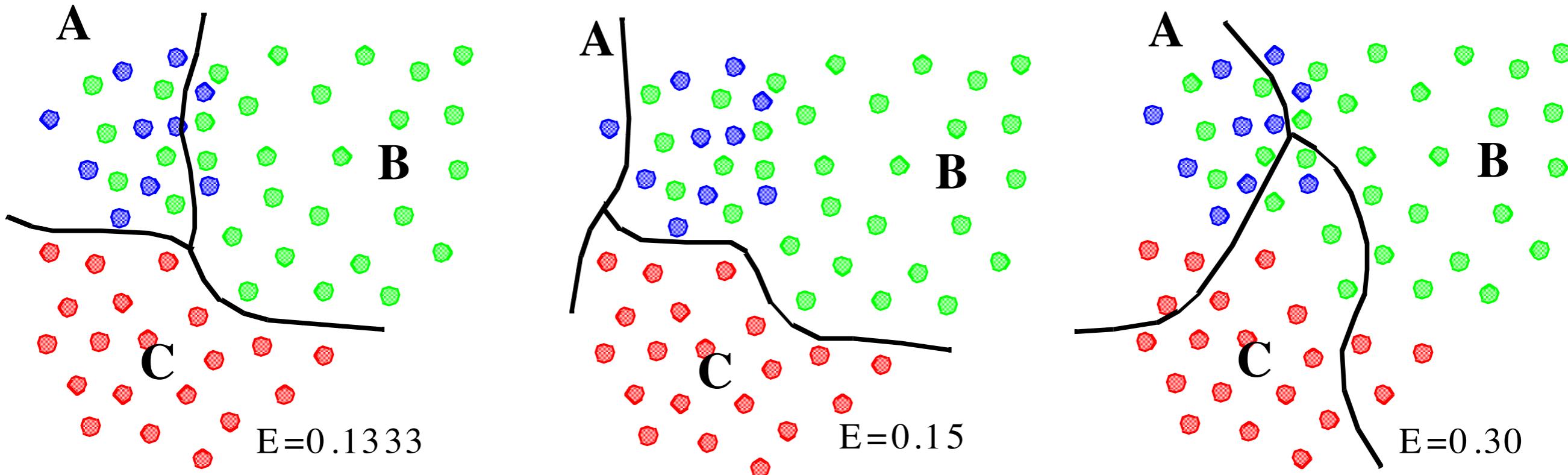
Λ real labels

L obtained labels

		classified to			
		A	B	C	
objects from	class A	8	2	0	0.20 error in class A
	class B	6	23	1	0.23 error in class B
	class C	4	1	15	0.25 error in class C
					0.228 averaged error



Confusion Matrix (3)



objects from

classified to

	A	B	C	
class A	8	2	0	0.20
class B	6	24	0	0.20
class C	0	0	20	0.00
	<hr/>			0.133

classified to

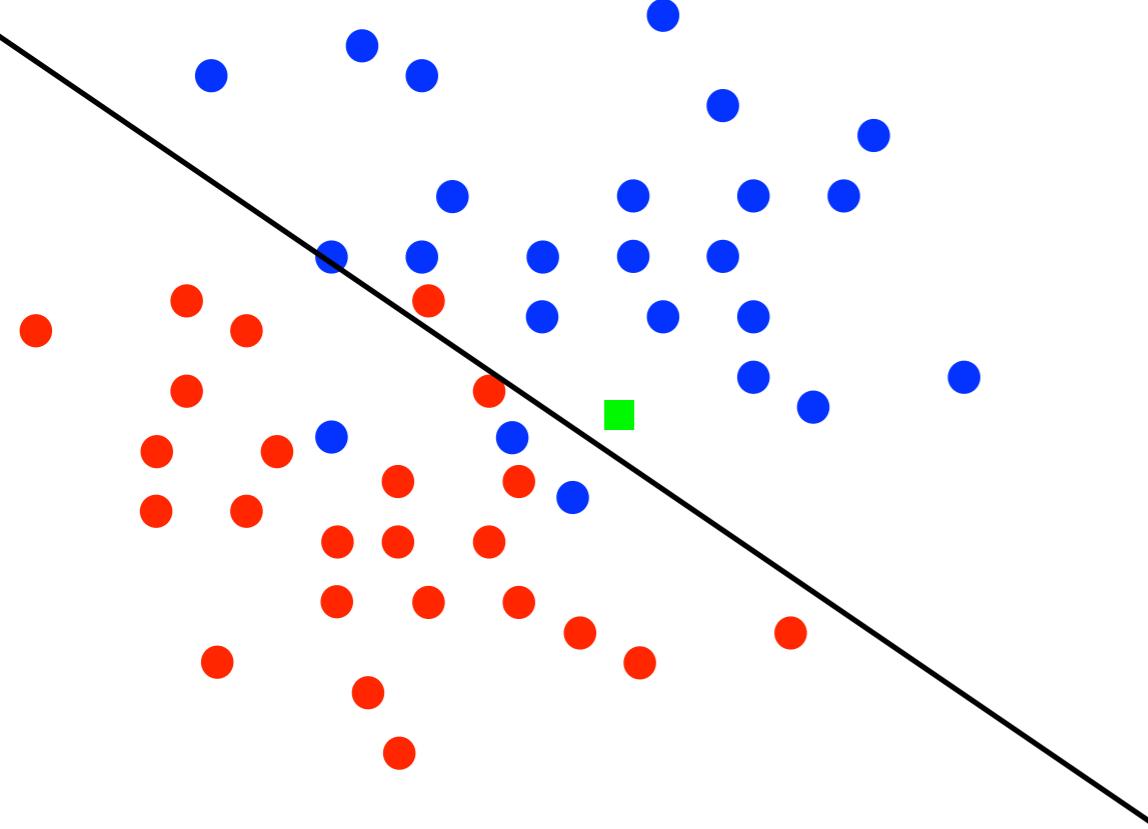
	A	B	C	
class A	1	9	0	0.9
class B	0	30	0	0.0
class C	0	0	20	0.0
	<hr/>			0.30

classified to

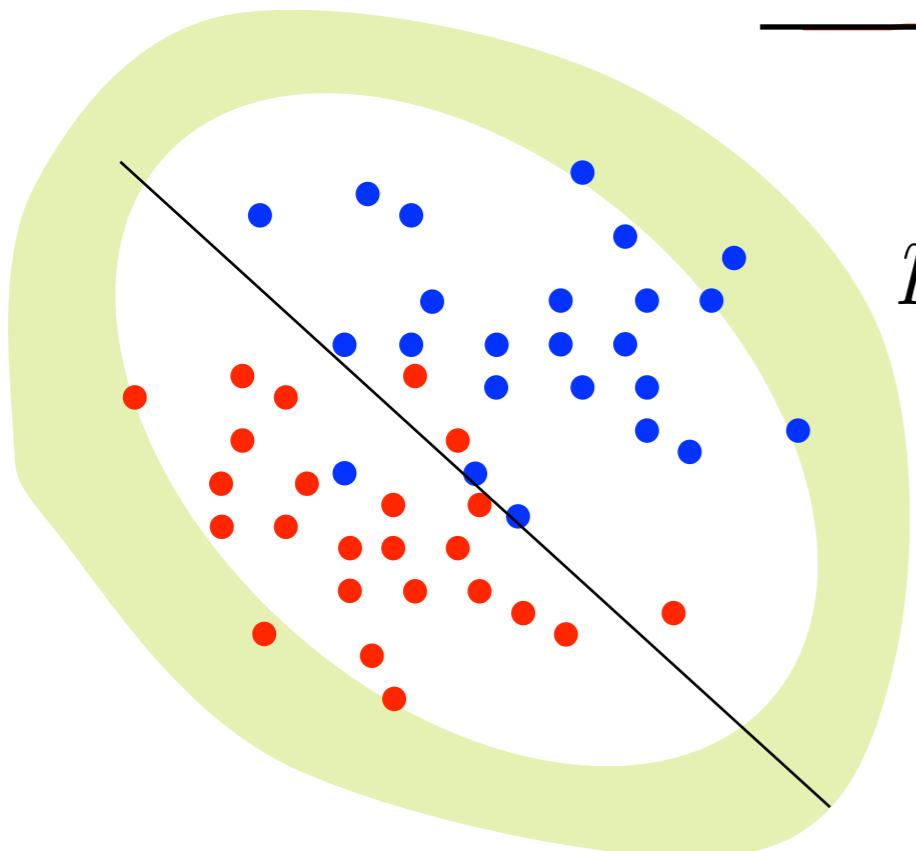
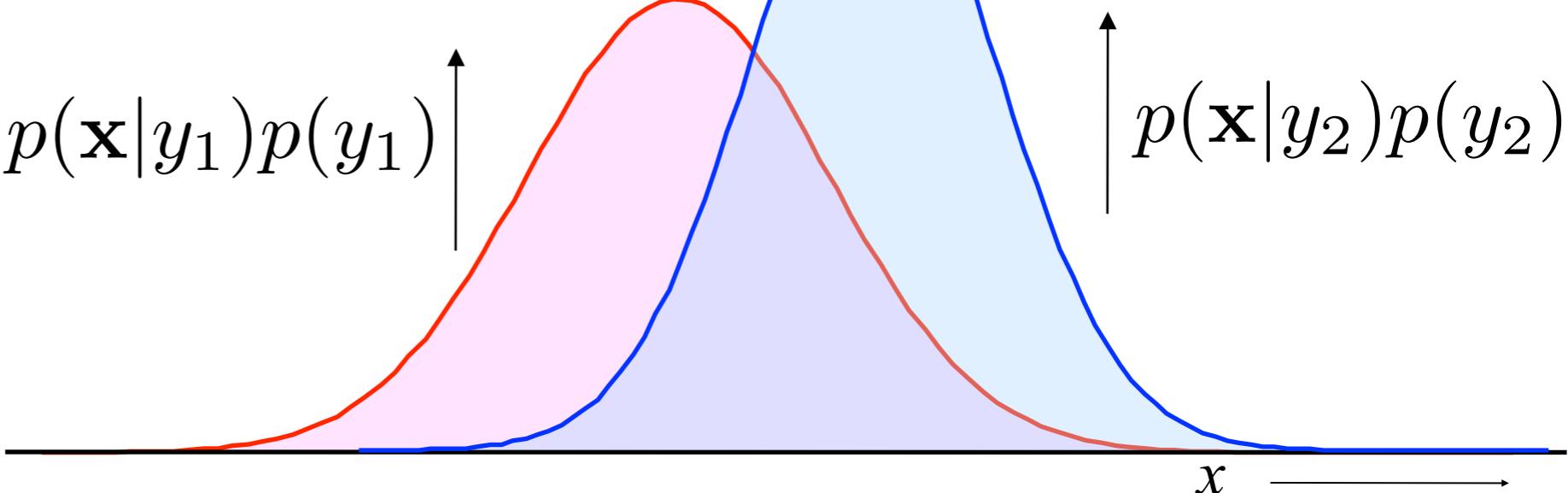
	A	B	C	
class A	7	2	1	0.30
class B	5	21	4	0.30
class C	3	2	15	0.30
	<hr/>			0.30

classification details are only observable in the confusion matrix!!

Labels of Squares? How Sure Are We?



Outlier Reject

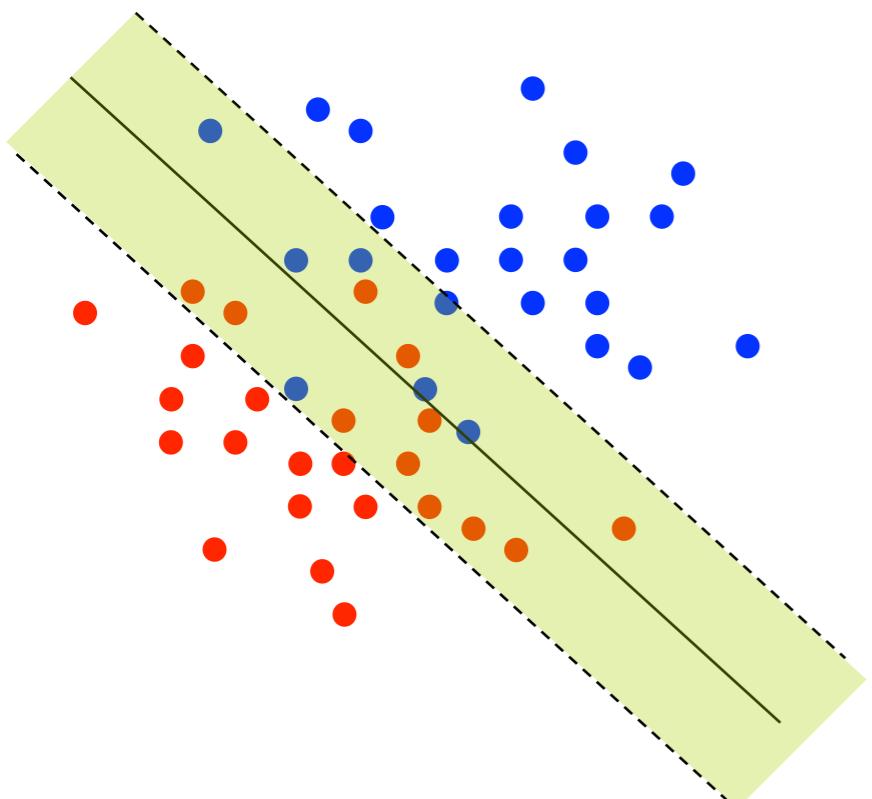


$$p(\mathbf{x}) = p(\mathbf{x}|y_1)p(y_1) + p(\mathbf{x}|y_2)p(y_2) \approx 0$$

$$|f(\mathbf{x})| \gg 0$$

Ambiguity Reject

- Reject objects for which classification is unsure, i.e., about equal posterior probabilities



$$p(y_1|\mathbf{x}) \approx p(y_2|\mathbf{x})$$

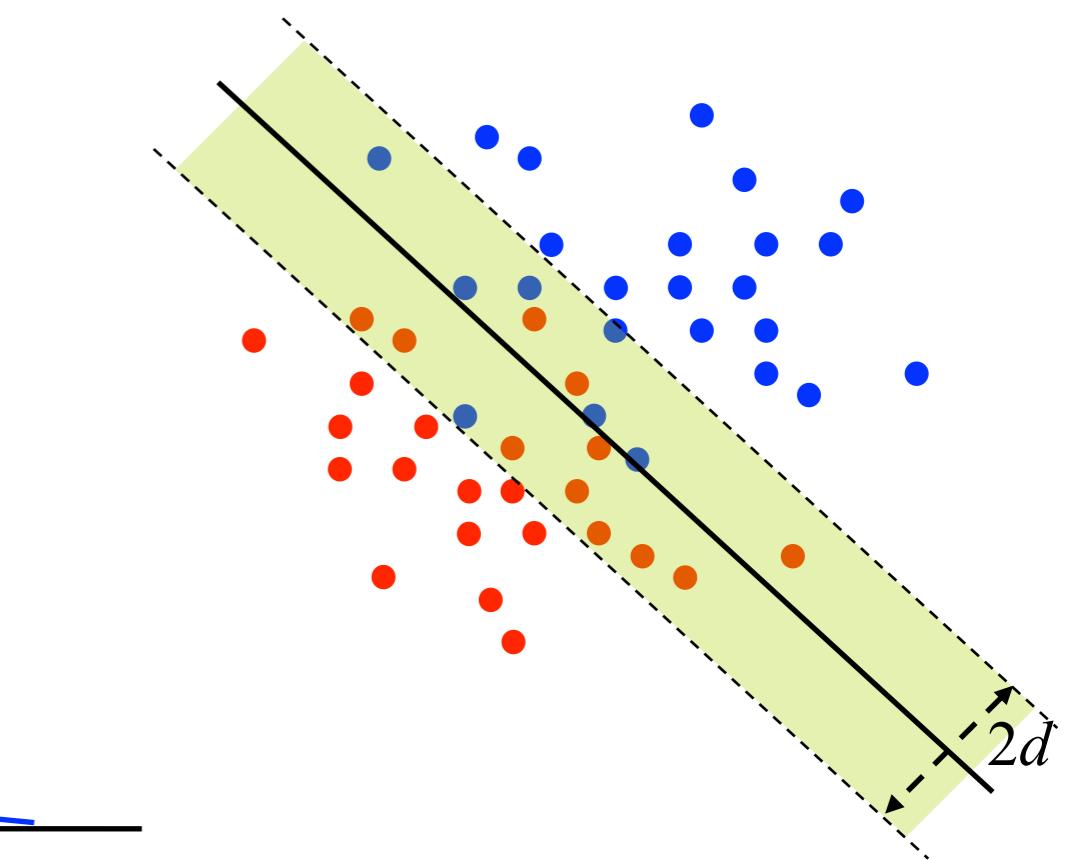
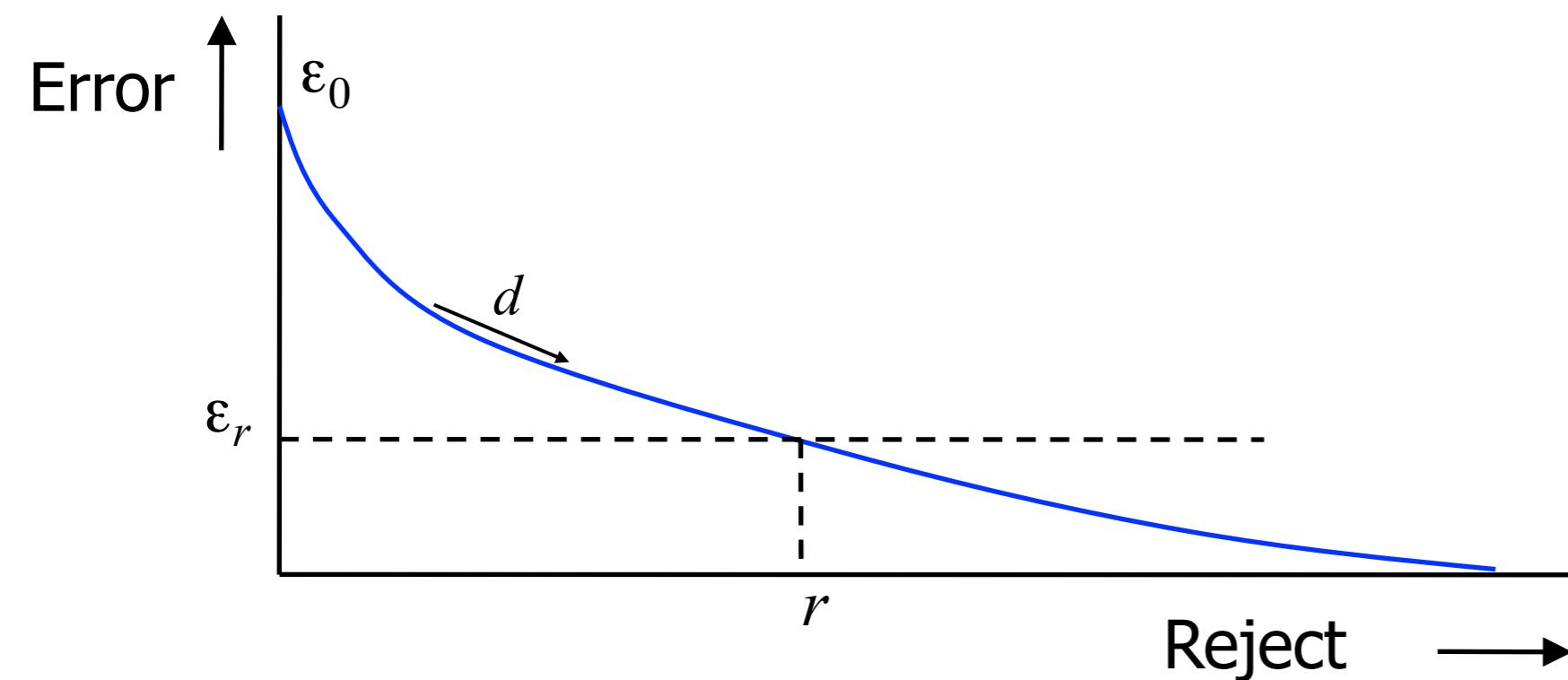
$$\frac{p(\mathbf{x}|y_1)p(y_1)}{p(\mathbf{x})} \approx \frac{p(\mathbf{x}|y_2)p(y_2)}{p(\mathbf{x})}$$

$$p(\mathbf{x}|y_1)p(y_1) - p(\mathbf{x}|y_2)p(y_2) \approx 0$$

$$f(\mathbf{x}) \approx 0$$

Reject Curve

- Classification error ε_0 can be reduced to ε_r by rejecting a fraction r of the objects
- N.B. Rejection also costs!



How much to reject?

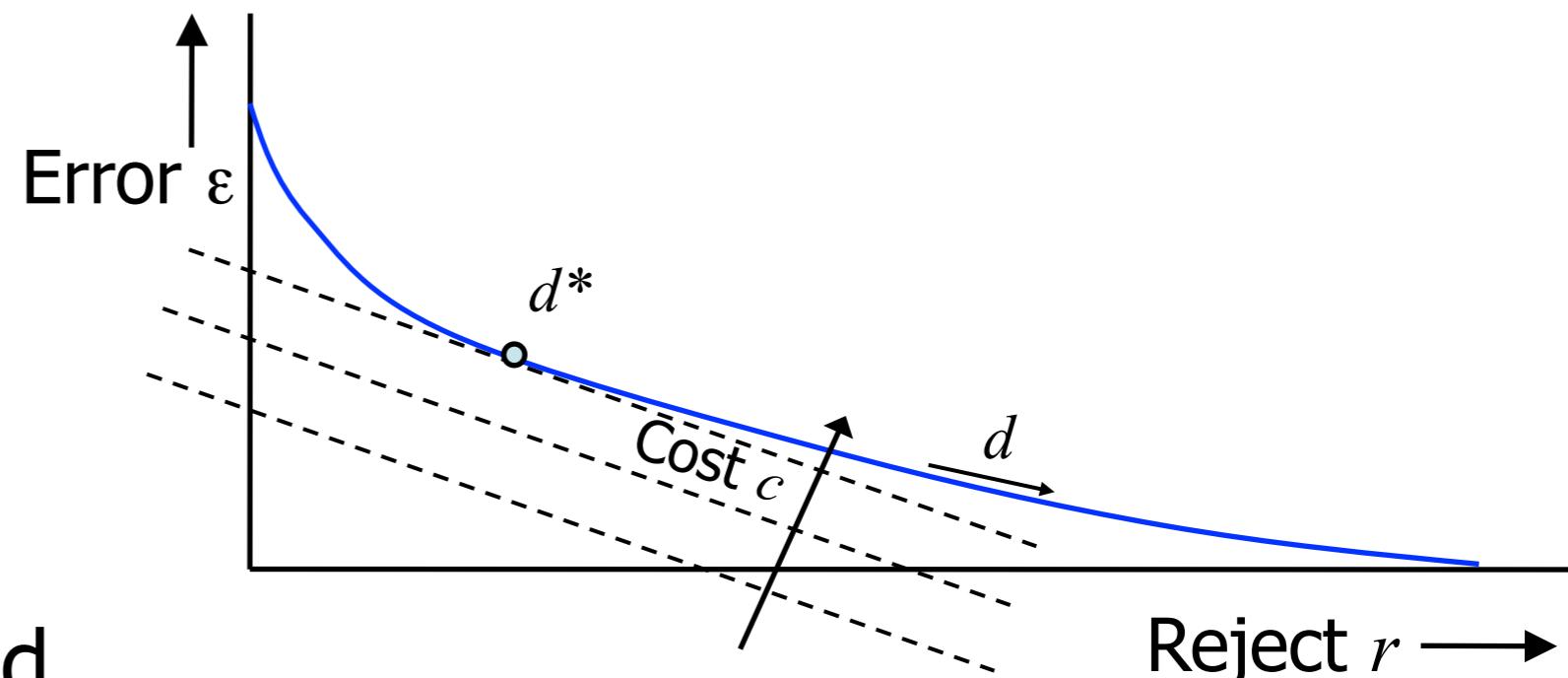
Compare the cost of a rejected object, c_r , with the cost of a classification error, c_ε :

$$c = c_r P(\text{reject}) + c_\varepsilon P(\text{error})$$

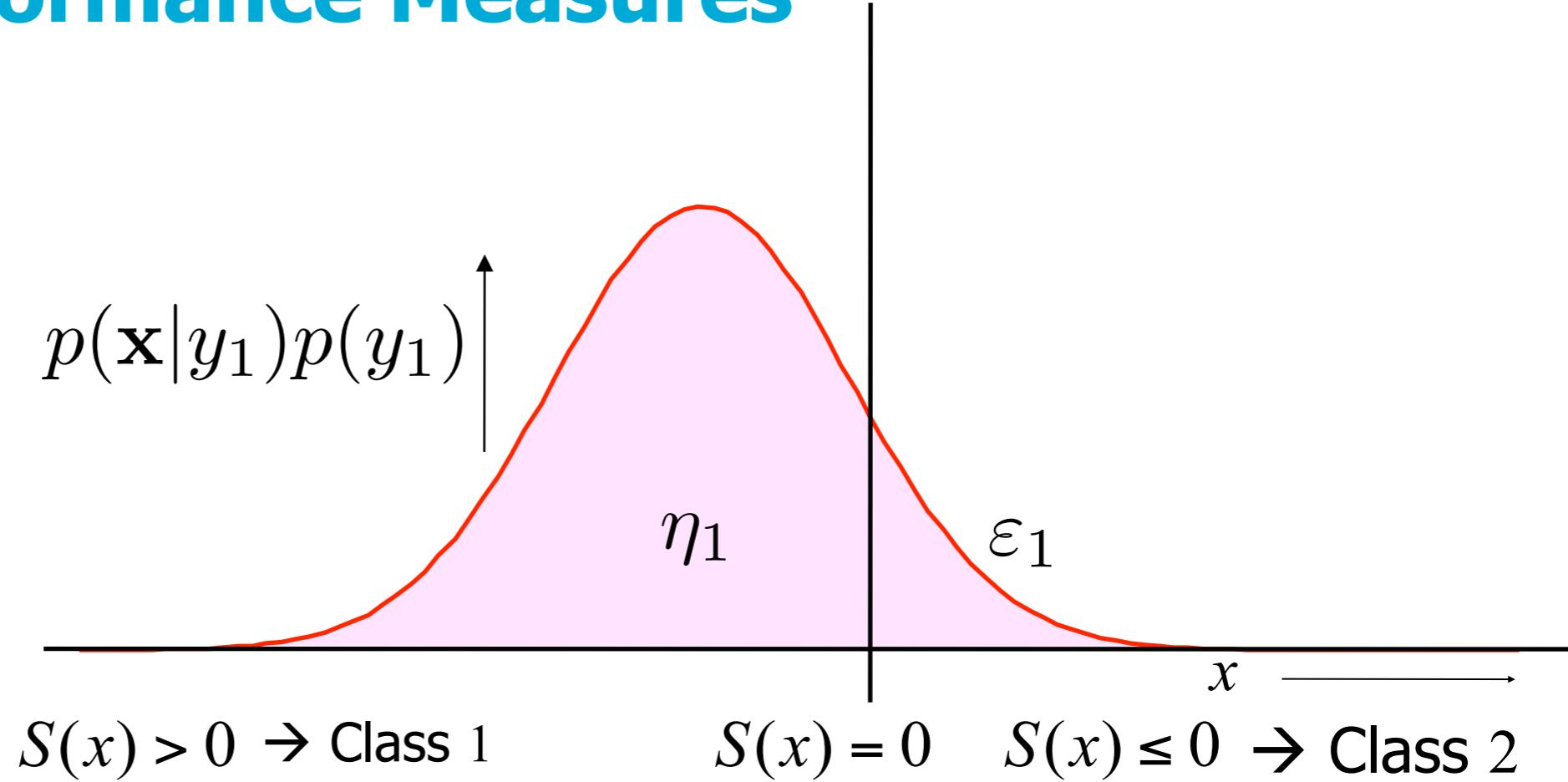
$$c = c_r r + c_\varepsilon \varepsilon$$

For given total cost c this is a linear function in the (r, ε) space.

Shift it until a possible operating point is reached.

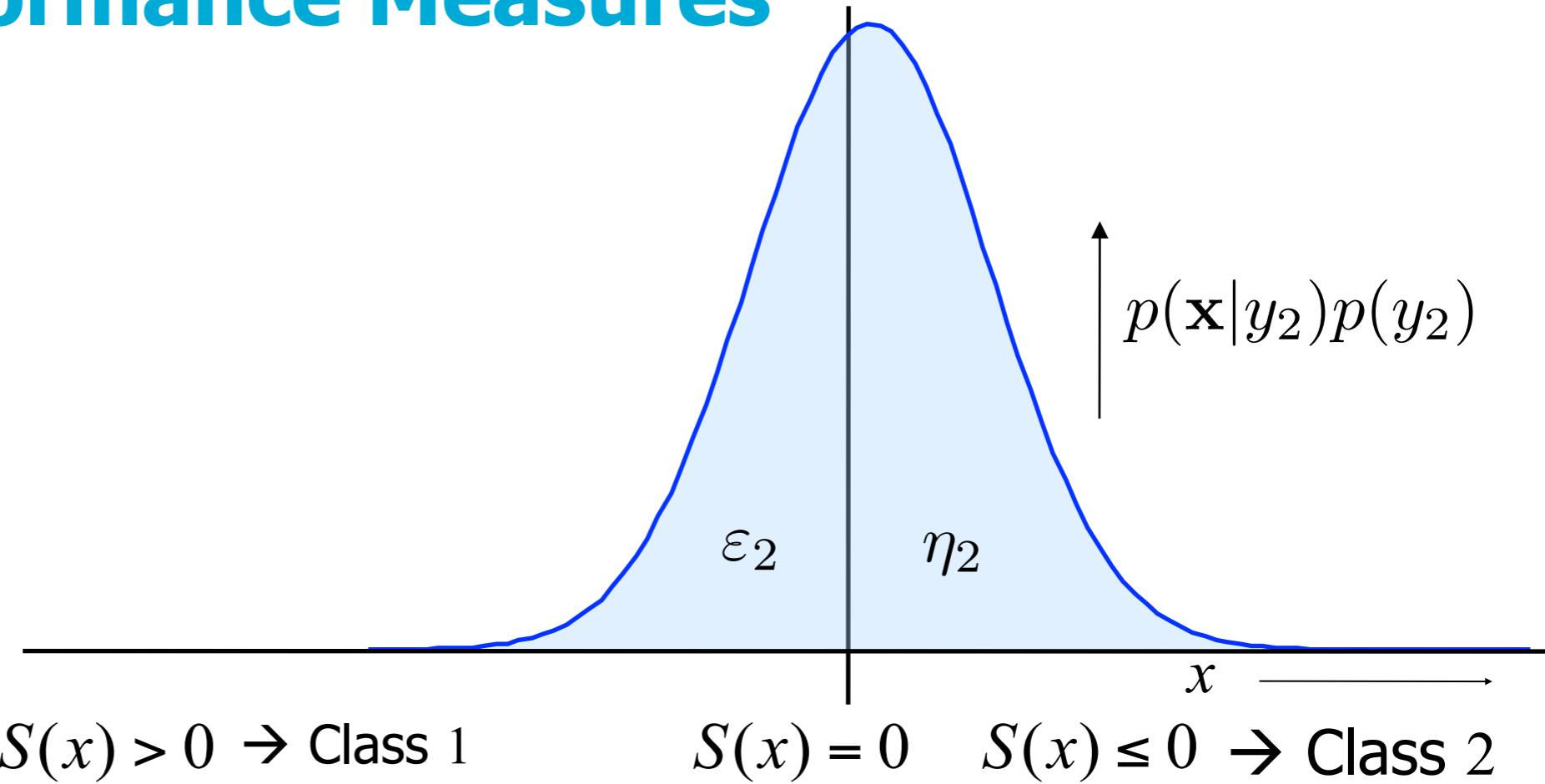


Error / Performance Measures



Objects of class 1, with prior $p(y_1)$, is classified by $S(x)$ in a part η_1 , assigned to y_1 , and a part ε_1 , assigned to y_2 .

Error / Performance Measures



Objects of class 2, with prior $p(y_2)$ is classified by $S(x)$ in a part η_2 , assigned to y_2 , and a part ε_2 , assigned to y_1 .

Error / Performance Measures

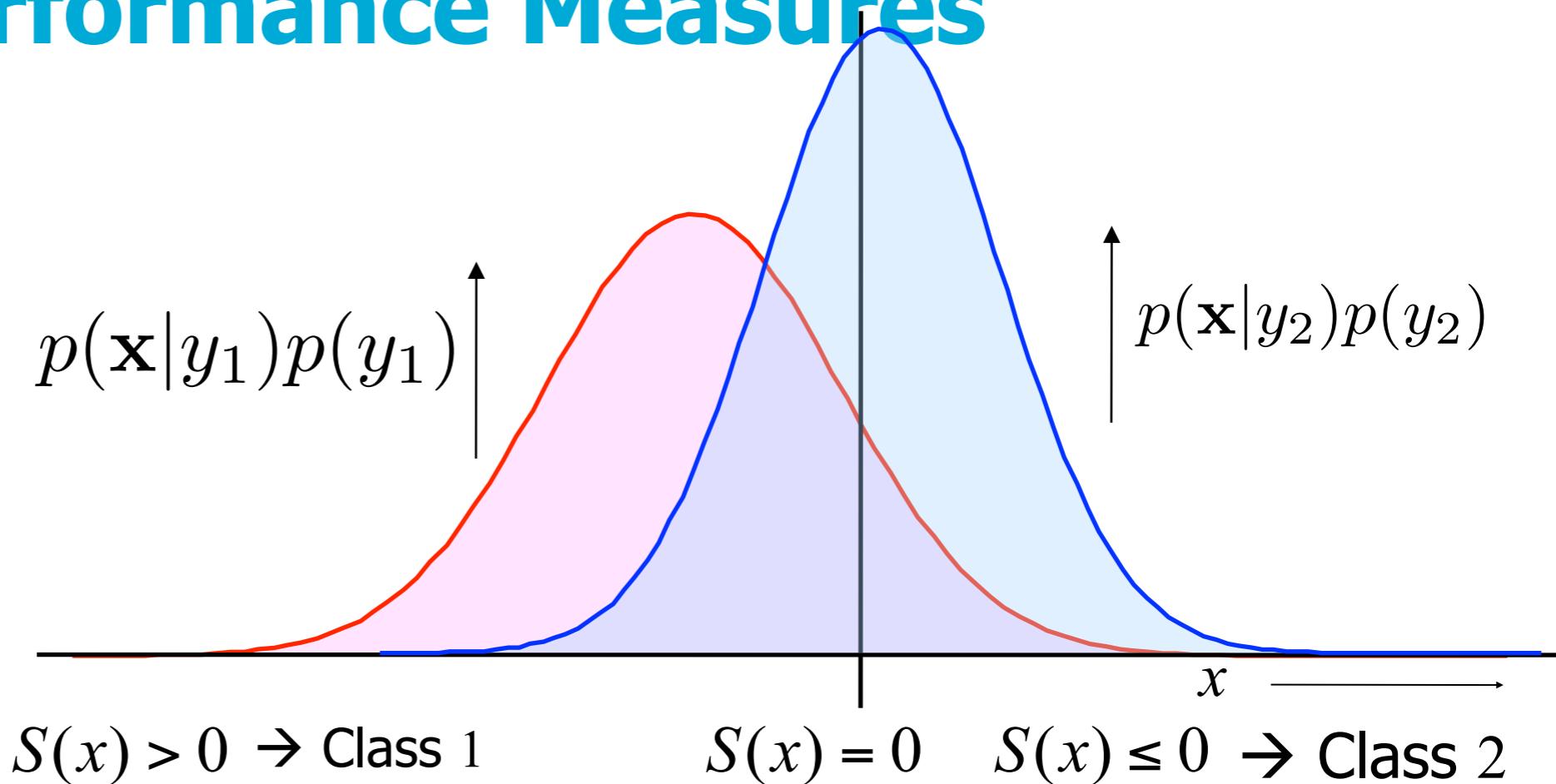
$$\varepsilon = \varepsilon_1 + \varepsilon_2$$

$$\eta = \eta_1 + \eta_2$$

$$p(y_1) + p(y_2) = 1$$

$$p(y_1) = \eta_1 + \varepsilon_1$$

$$p(y_2) = \eta_2 + \varepsilon_2$$



ε_1 : error contribution class A

ε_2 : error contribution class B

ε : classification error

η : performance / accuracy

$$\varepsilon + \eta = 1$$

$\varepsilon_A/p(y_1)$: error in class 1

$\varepsilon_B/p(y_2)$: error in class 2

$\eta_1/p(y_1)$: sensitivity [recall] class 1

[what goes correct in class 1]

$\eta_1/(\eta_1 + \varepsilon_2)$: precision class 1

[what belongs to A in what is classified as A]

Error / Performance Measures

- **Error** : probability of erroneous classifications
- **Performance / accuracy** : $1 - \text{error}$
- **Sensitivity** of a target class [e.g. diseased patients] : performance for objects from that target class
- **Specificity** : performance for all objects outside target class
- **Precision** of a target class : fraction of correct objects among all objects assigned to that class.
- **Recall** : fraction of correctly classified objects; identical to sensitivity when related to particular class
- **True positive rate** : identical to sensitivity
- **False positive rate** : error for all objects outside target

Combining the two errors

- There is a fundamental tradeoff between the two errors/performances of the two classes

- Standard classification error:

$$\varepsilon = \varepsilon_1 p(y_1) + \varepsilon_2 p(y_2)$$

- Weighted classification error:

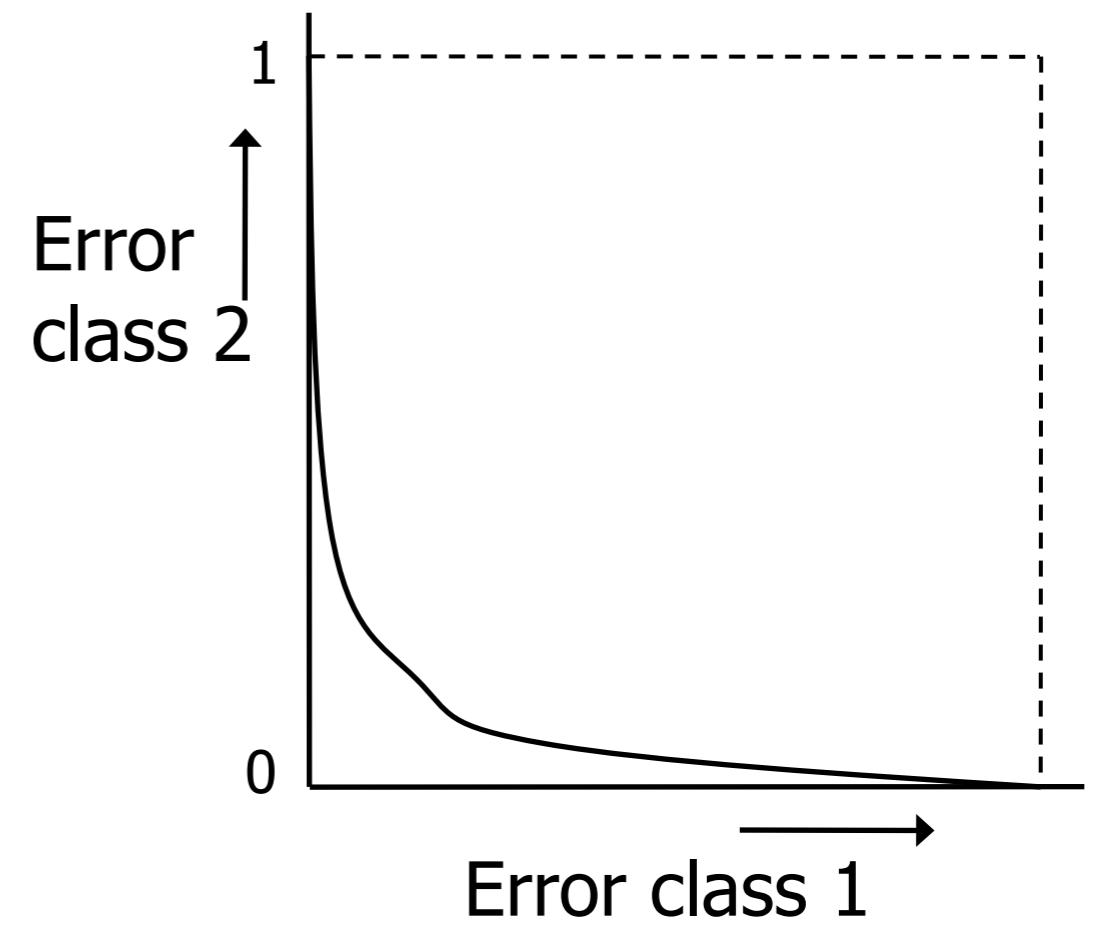
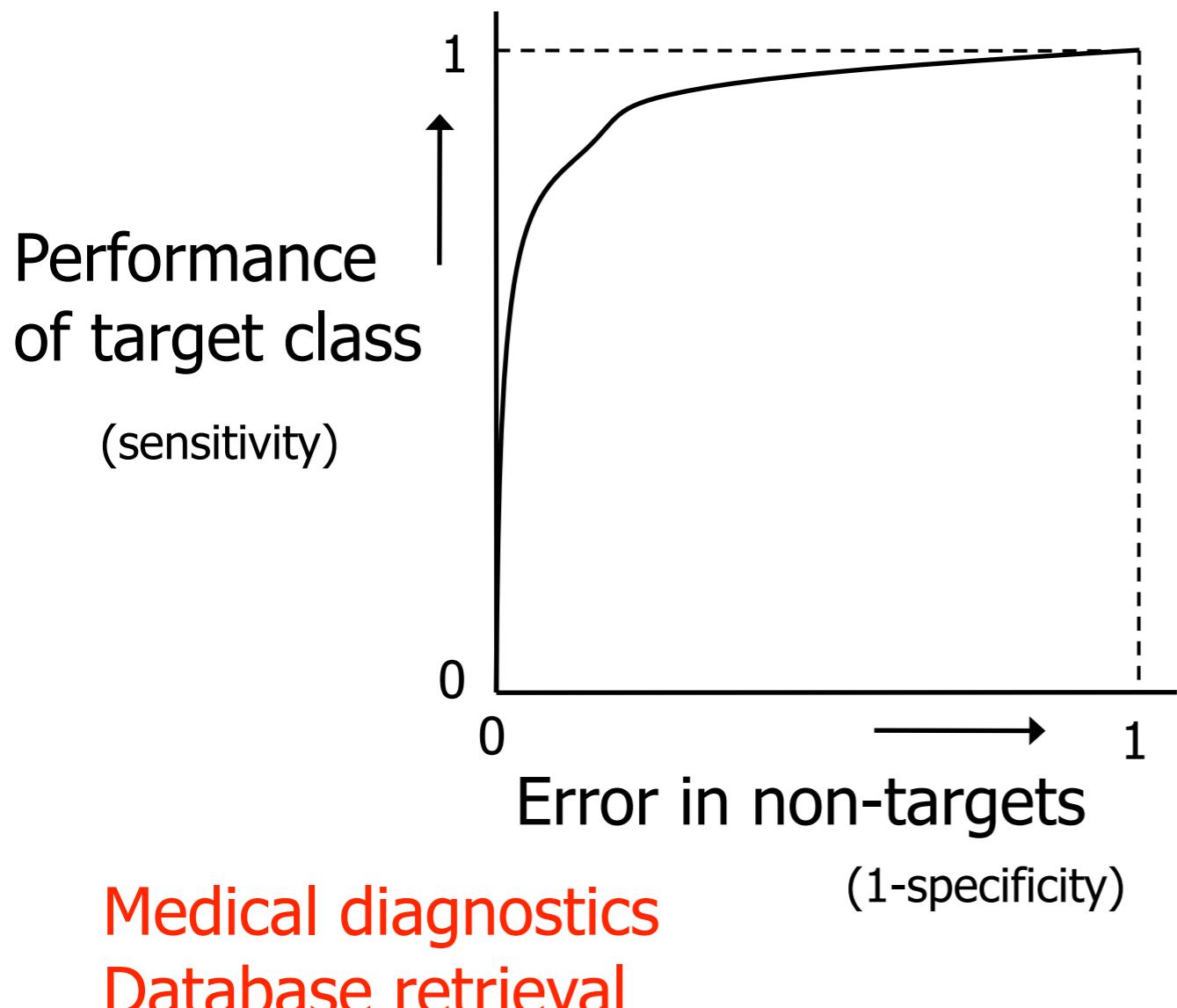
$$\varepsilon = \lambda_{12}\varepsilon_1 p(y_1) + \lambda_{21}\varepsilon_2 p(y_2)$$

- F1-score (harmonic mean):

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

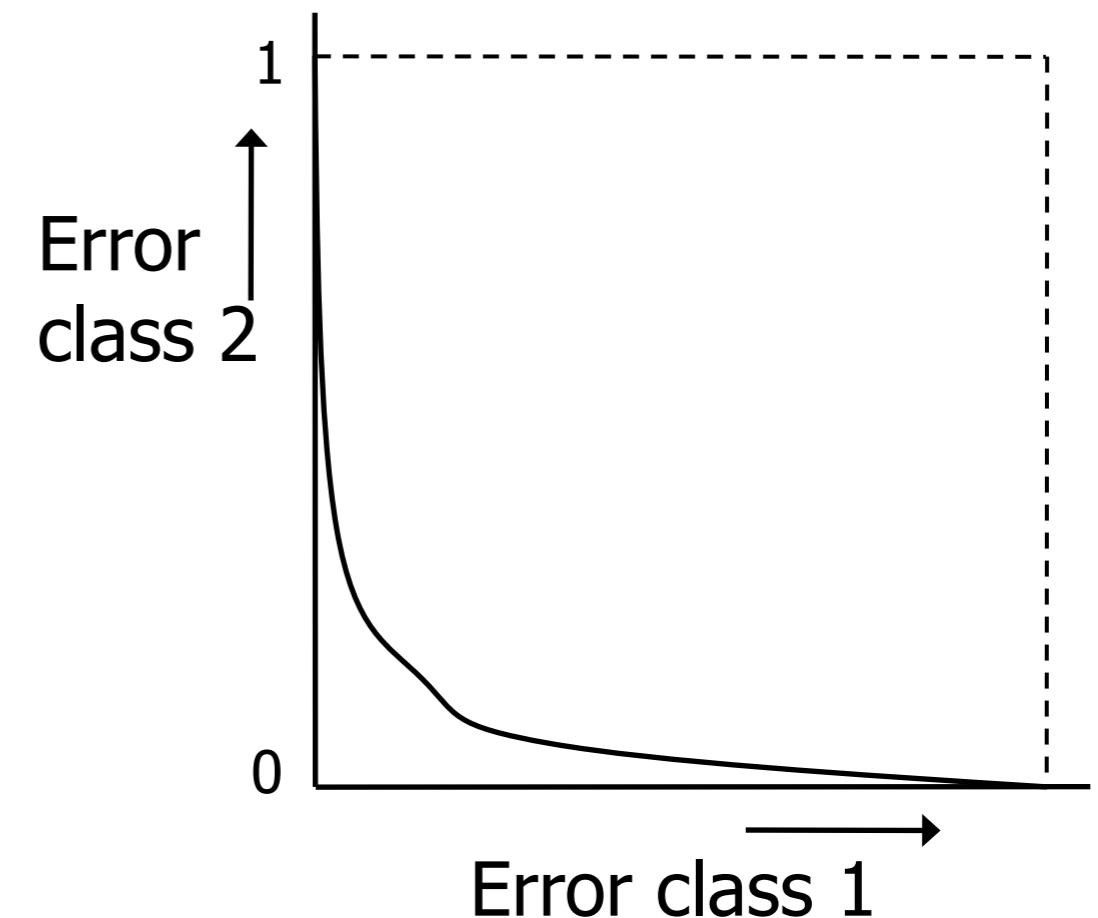
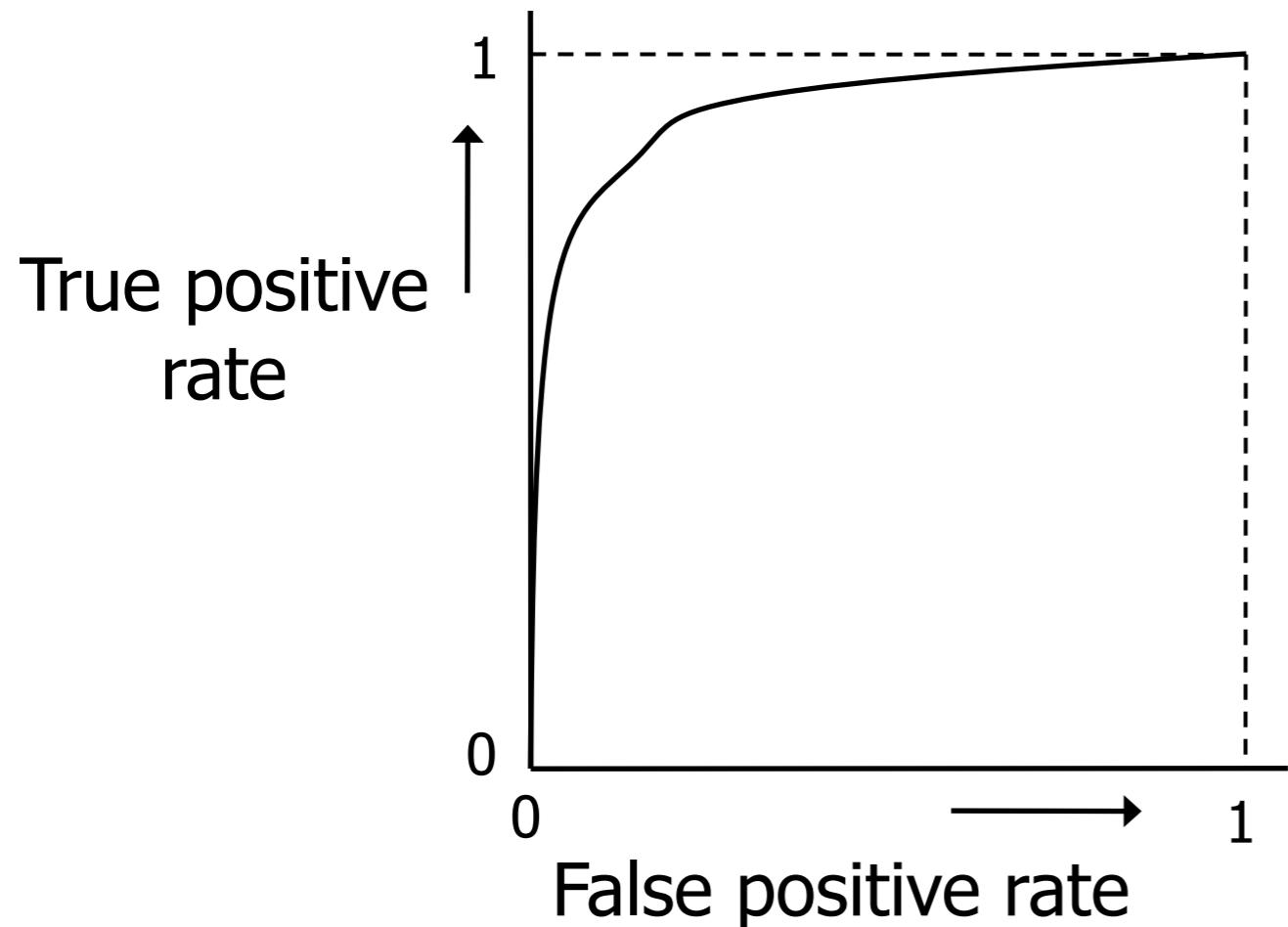
ROC Analysis

ROC: Receiver-Operator Characteristic (from communication theory)



ROC Analysis

ROC: Receiver-Operator Characteristic (from communication theory)

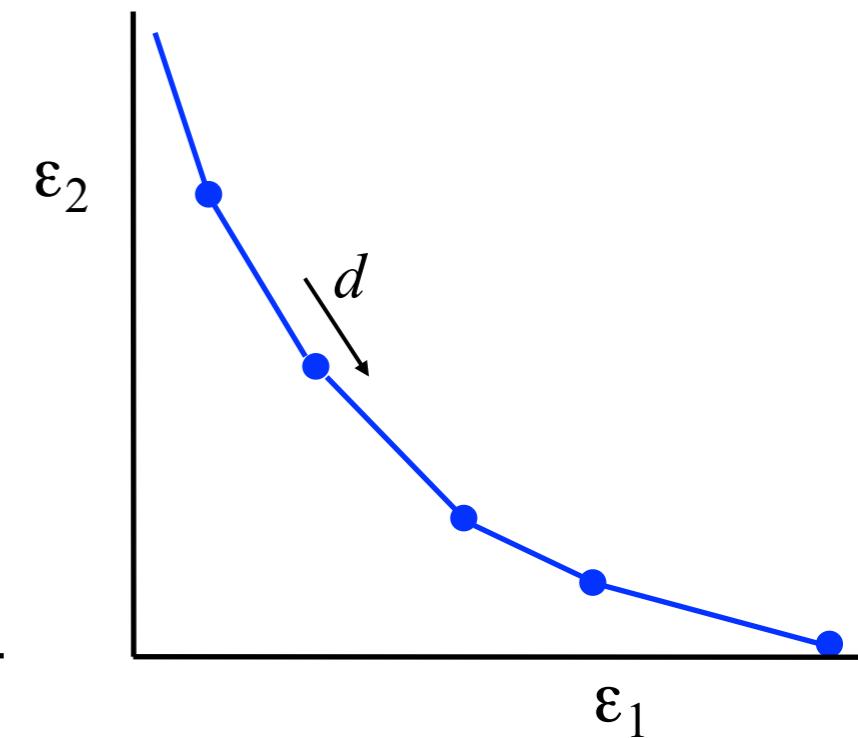
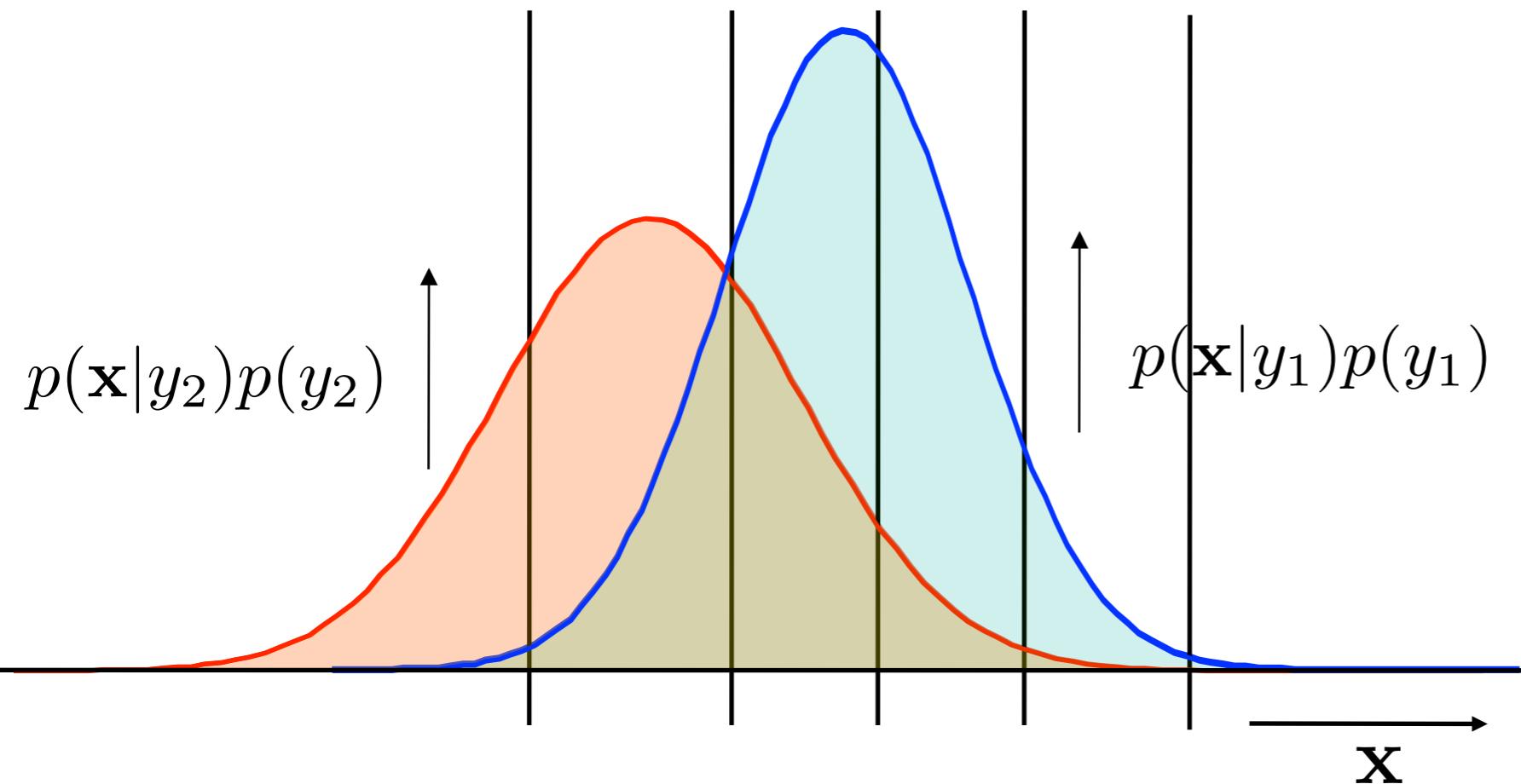


Medical diagnostics
Database retrieval

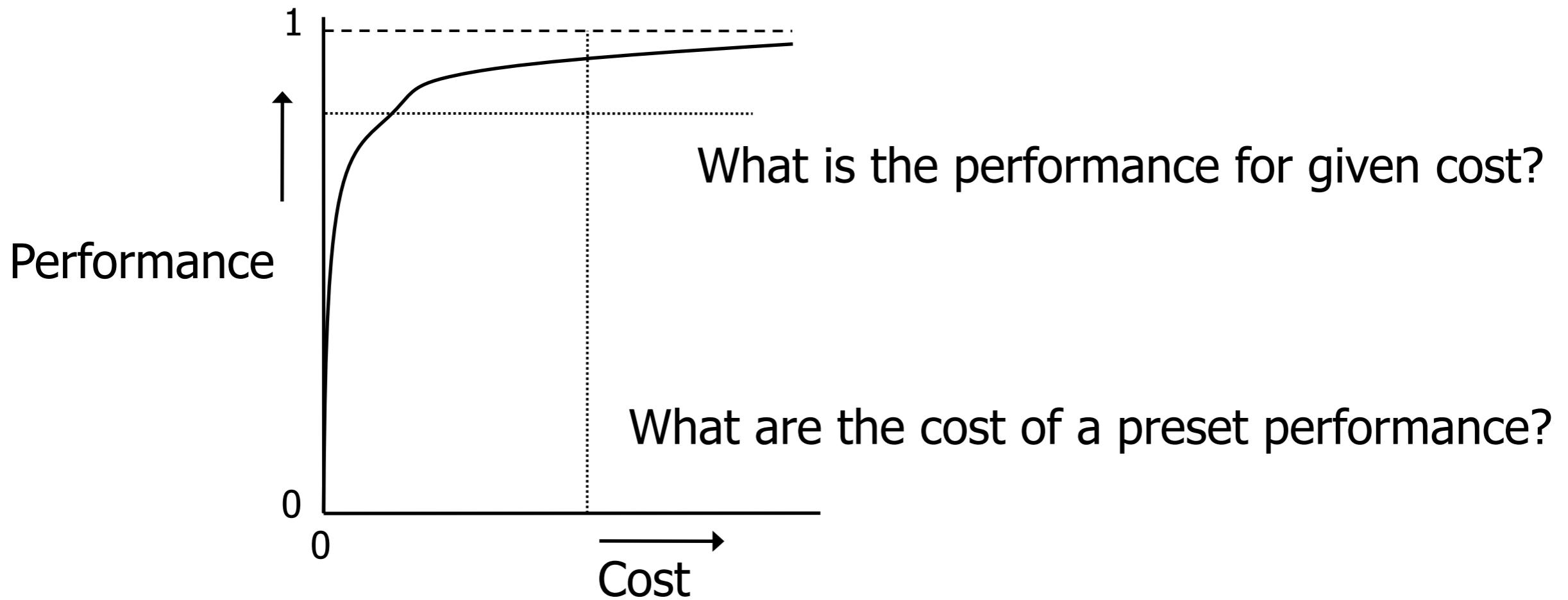
2-class pattern recognition

ROC Curve

- Curve is obtained by varying classifier threshold d

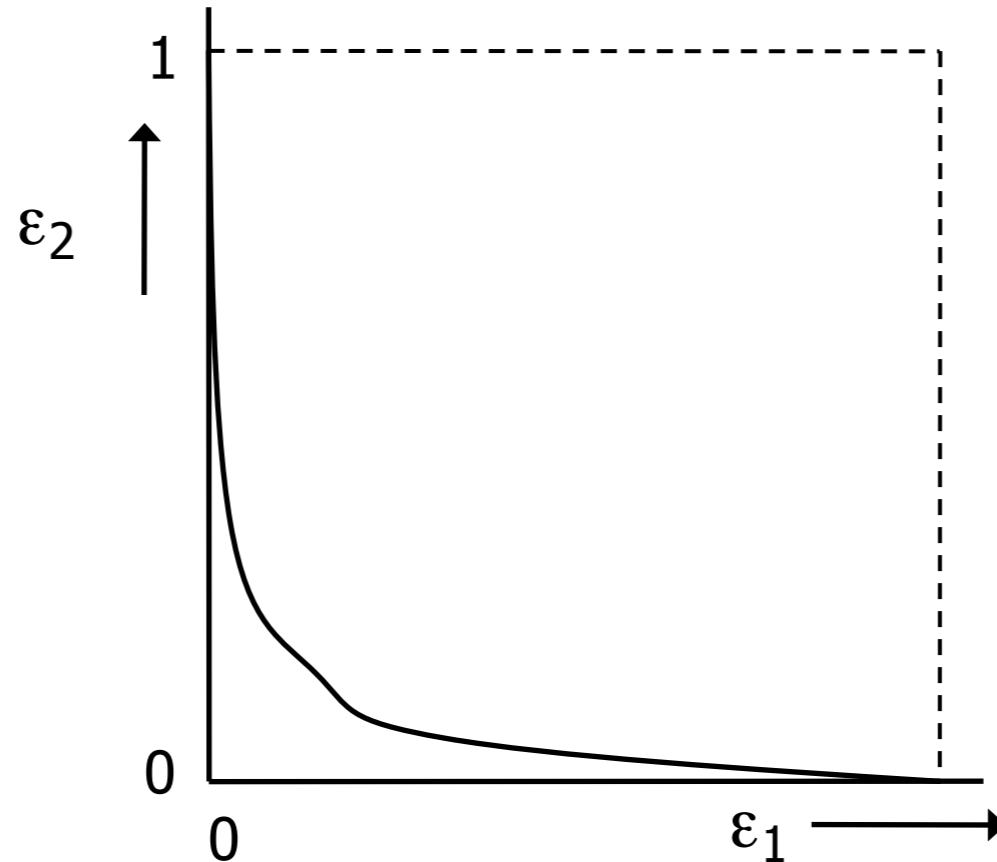


When are ROC Curves Useful?



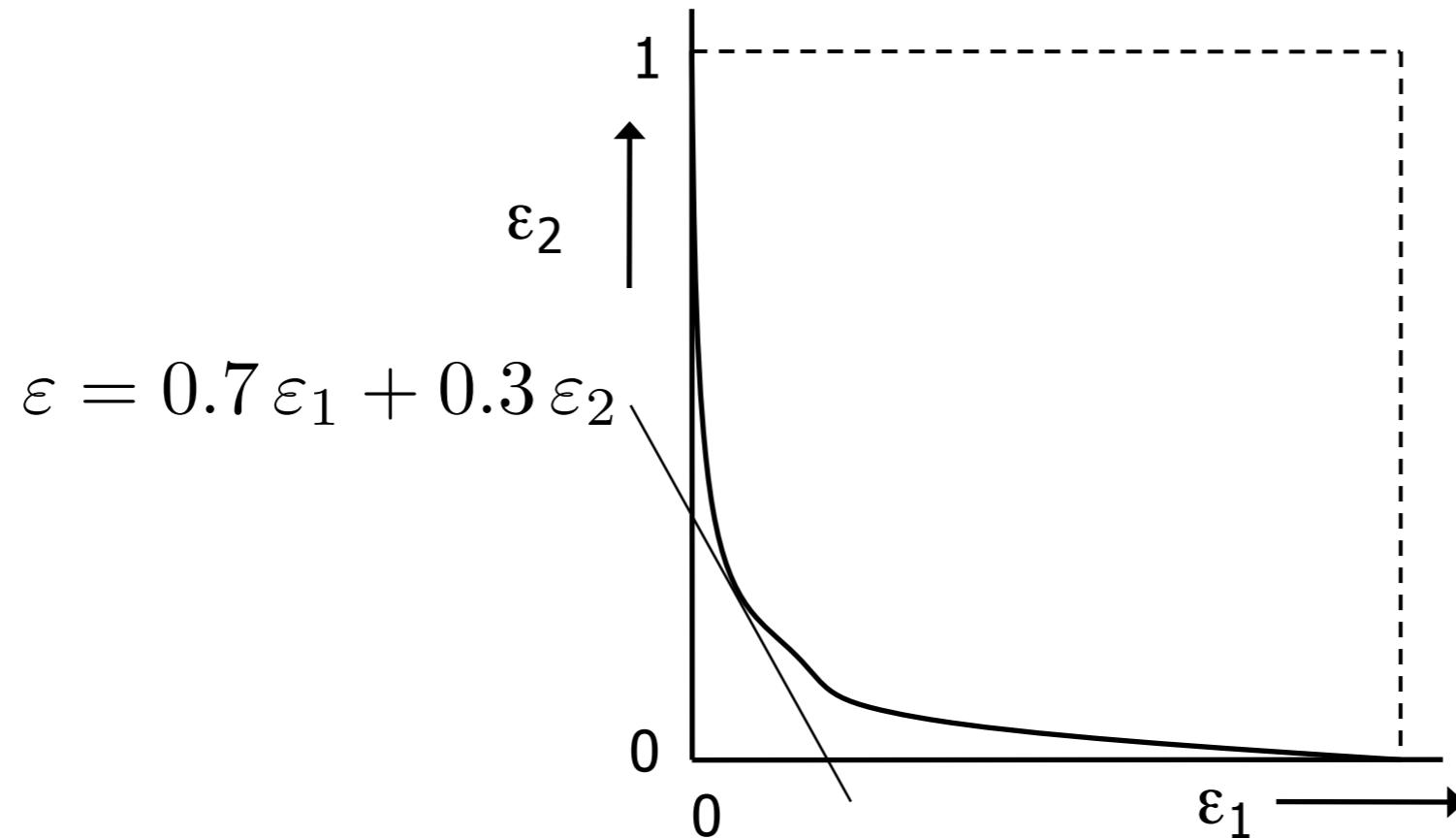
Trade-off between cost and performance.

Studying Effect of Changing Priors



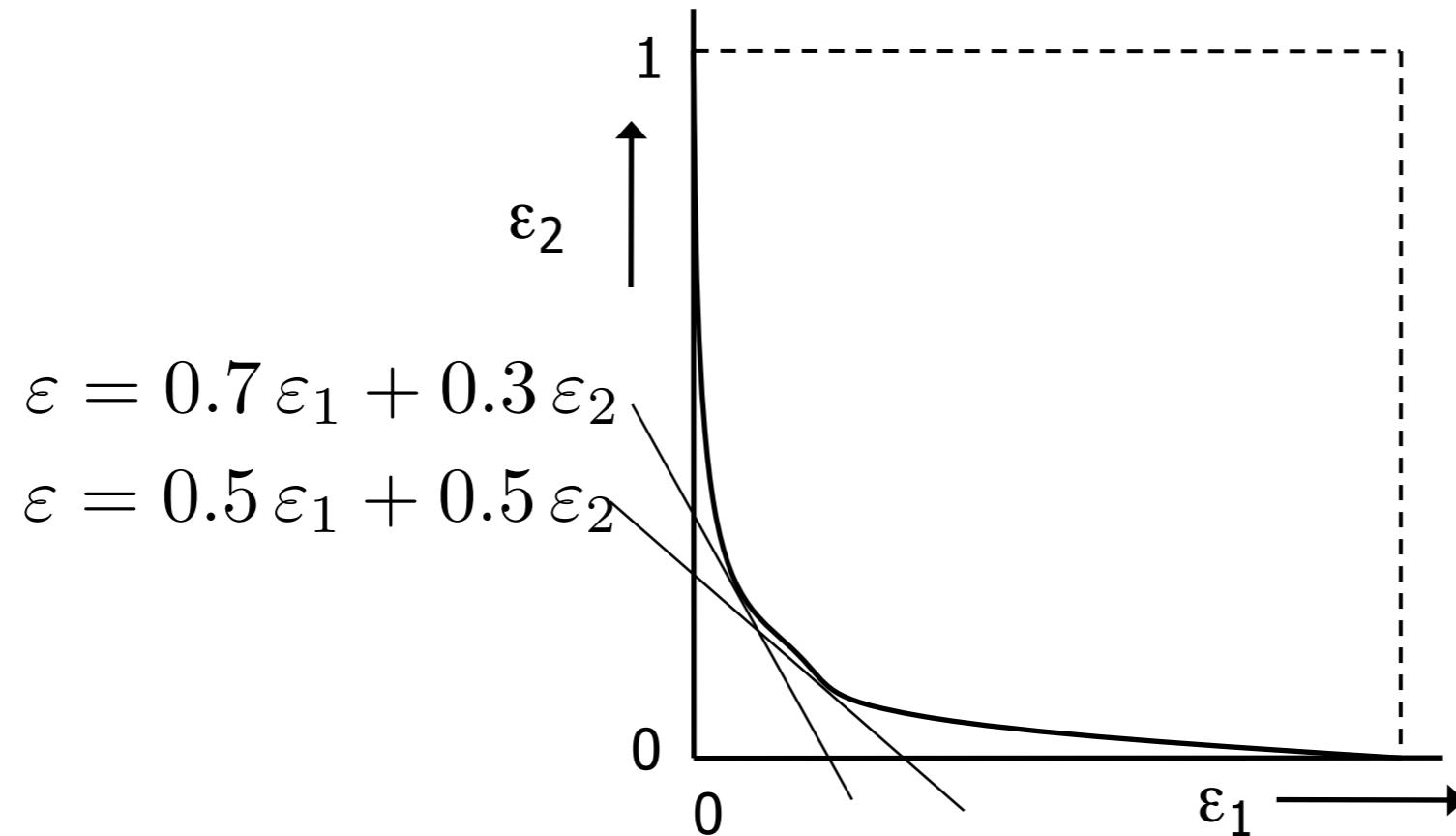
$$\varepsilon = p(y_1) \varepsilon_1 + p(y_2) \varepsilon_2$$

Studying Effect of Changing Priors



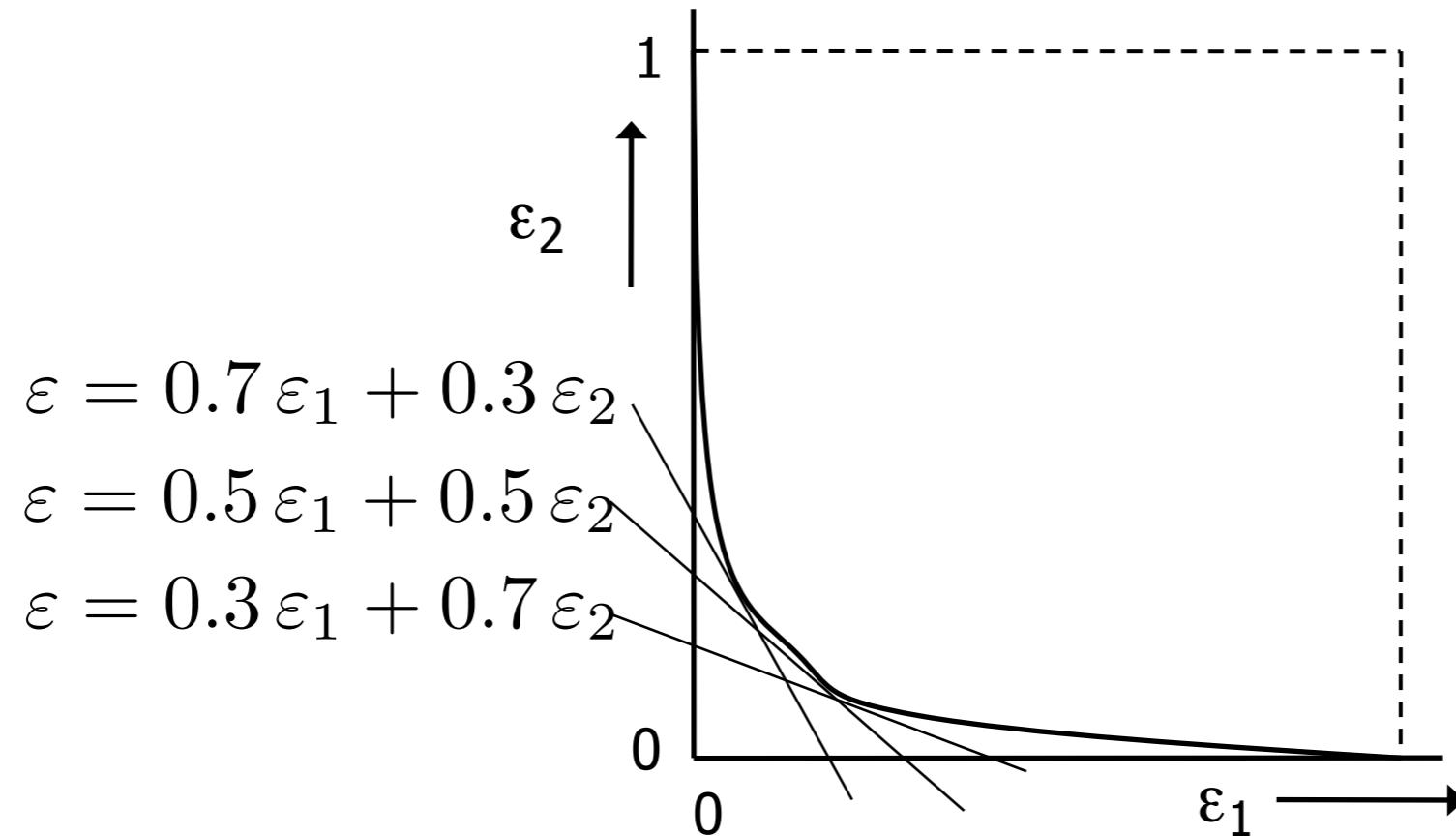
$$\varepsilon = p(y_1) \varepsilon_1 + p(y_2) \varepsilon_2$$

Studying Effect of Changing Priors



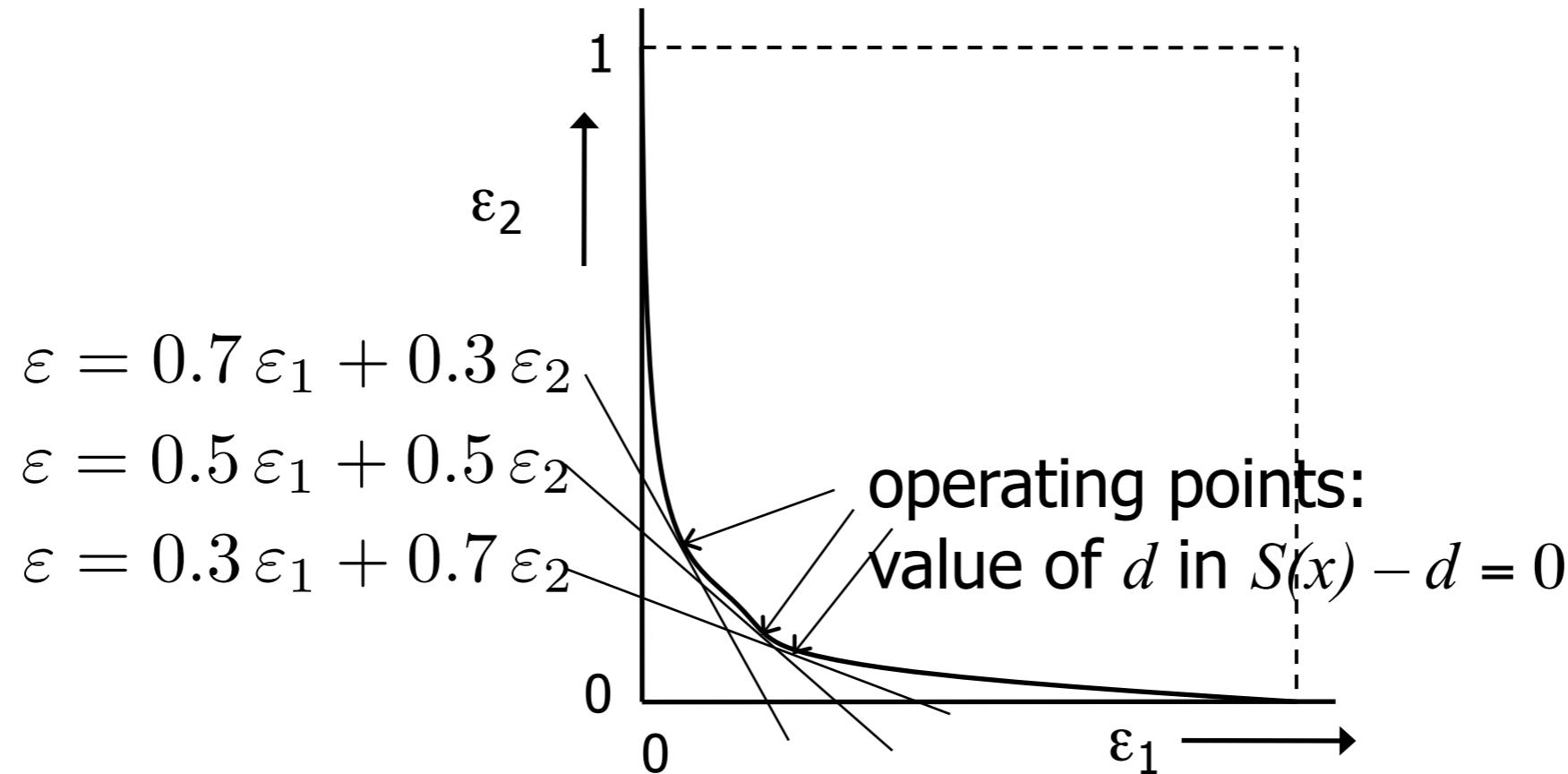
$$\varepsilon = p(y_1) \varepsilon_1 + p(y_2) \varepsilon_2$$

Studying Effect of Changing Priors



$$\varepsilon = p(y_1) \varepsilon_1 + p(y_2) \varepsilon_2$$

Studying Effect of Changing Priors



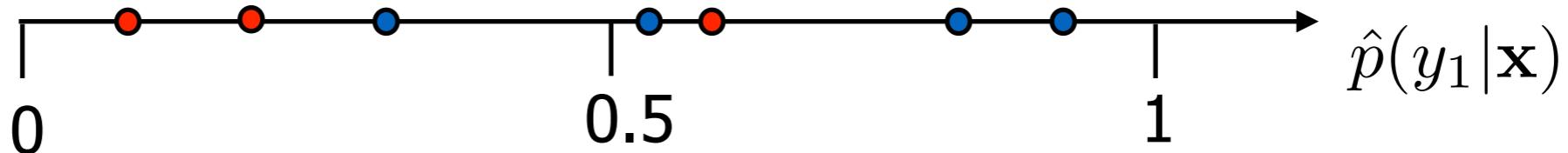
$$\varepsilon = p(y_1) \varepsilon_1 + p(y_2) \varepsilon_2$$

Example to compute ROC curve

- Assume I trained a classifier, and testing it on a test set:
- Four objects of class 1:
$$\hat{p}(y_1|\mathbf{x}_1) = 0.3$$
$$\hat{p}(y_1|\mathbf{x}_2) = 0.8$$
$$\hat{p}(y_1|\mathbf{x}_3) = 0.9$$
$$\hat{p}(y_1|\mathbf{x}_4) = 0.55$$
- Three objects of class 2:
$$\hat{p}(y_1|\mathbf{x}_5) = 0.2$$
$$\hat{p}(y_1|\mathbf{x}_6) = 0.1$$
$$\hat{p}(y_1|\mathbf{x}_7) = 0.6$$
- How does the ROC curve look like? ε_2 vs. ε_1

Example to compute ROC curve

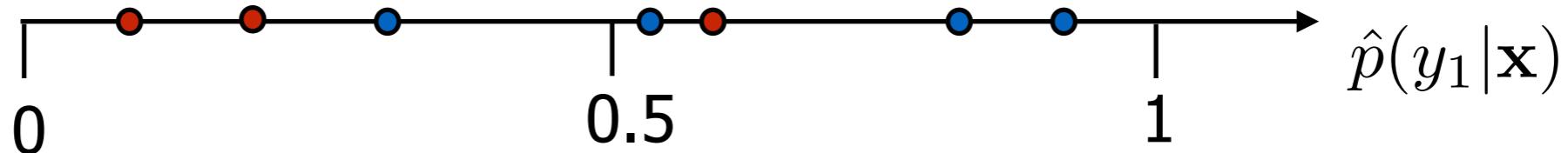
- Find out how obj's are classified for different thresholds



- For example, what are the errors for $d = 0.5$?

Example to compute ROC curve

- Find out how obj's are classified for different thresholds

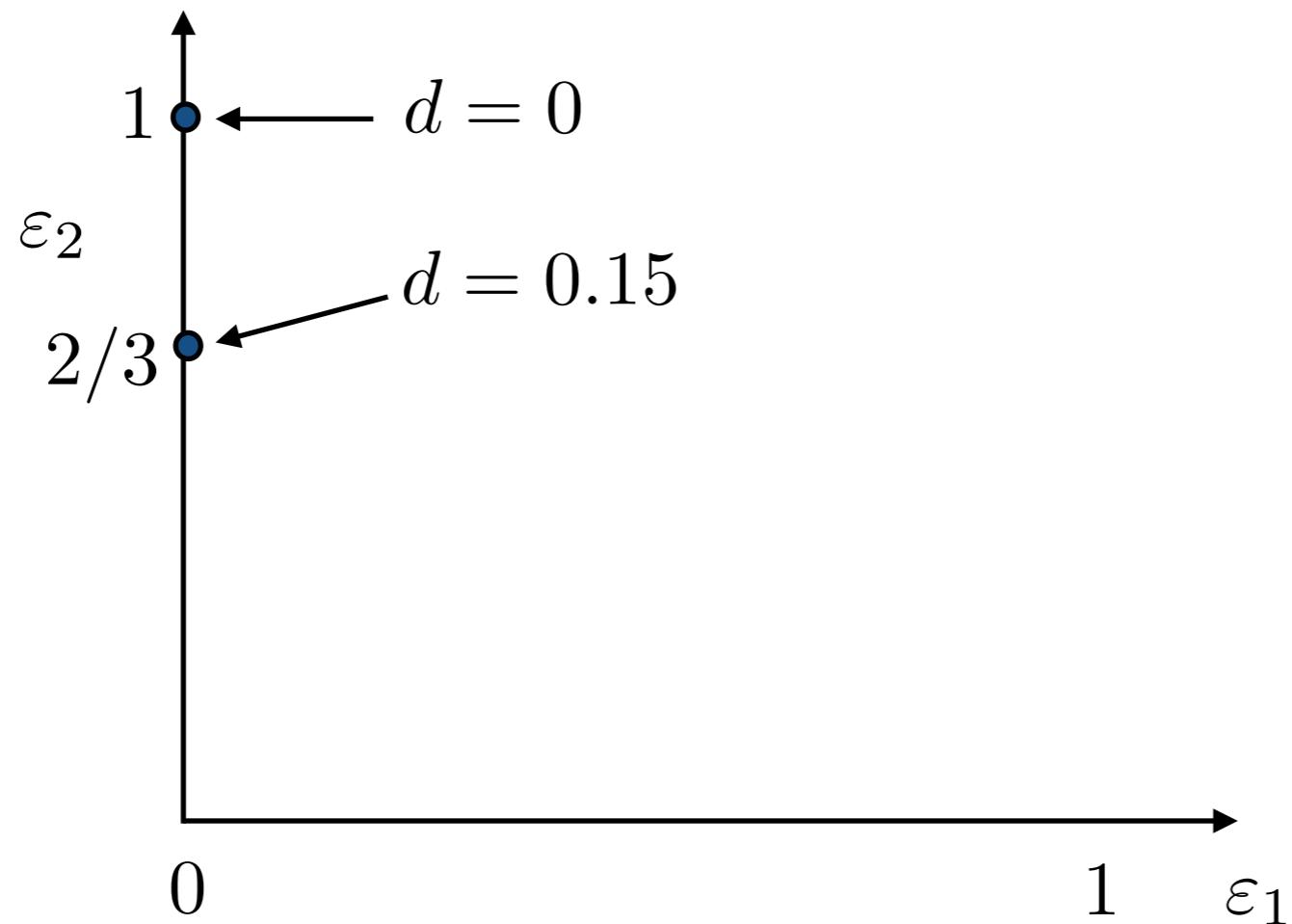


- For example, what are the errors for $d = 0.5$?

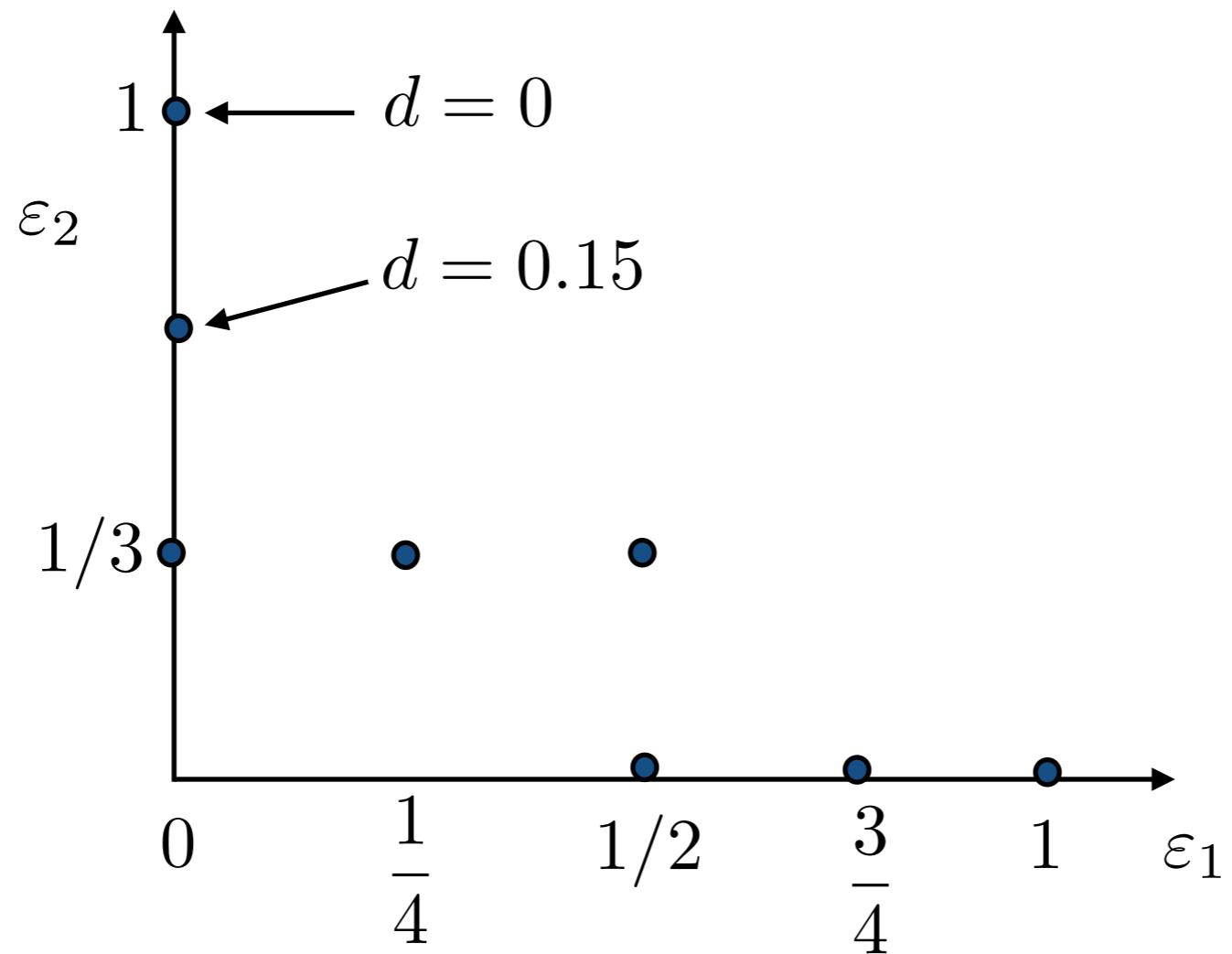
$$\varepsilon_1 = 1/4$$

$$\varepsilon_2 = 1/3$$

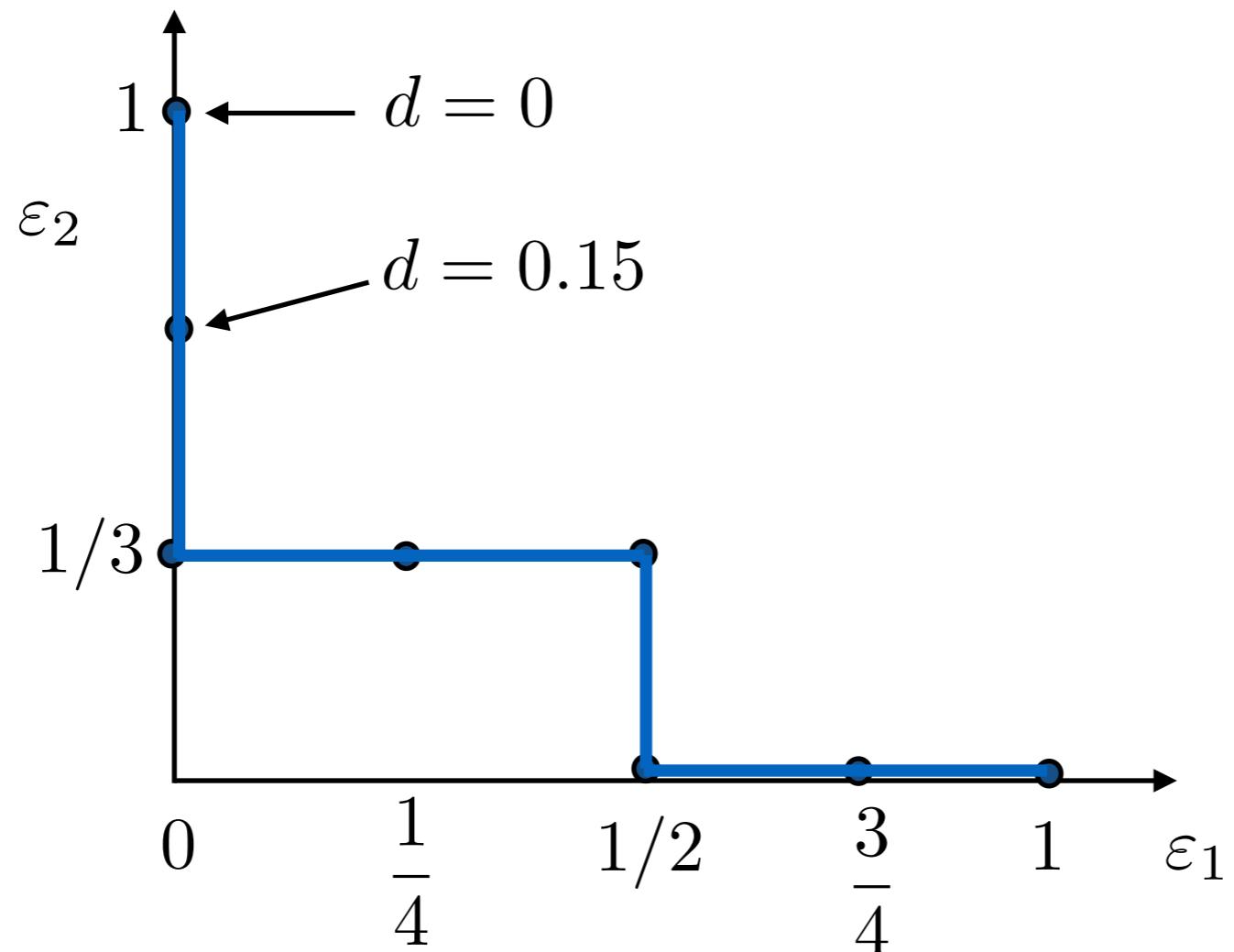
Example to compute ROC curve



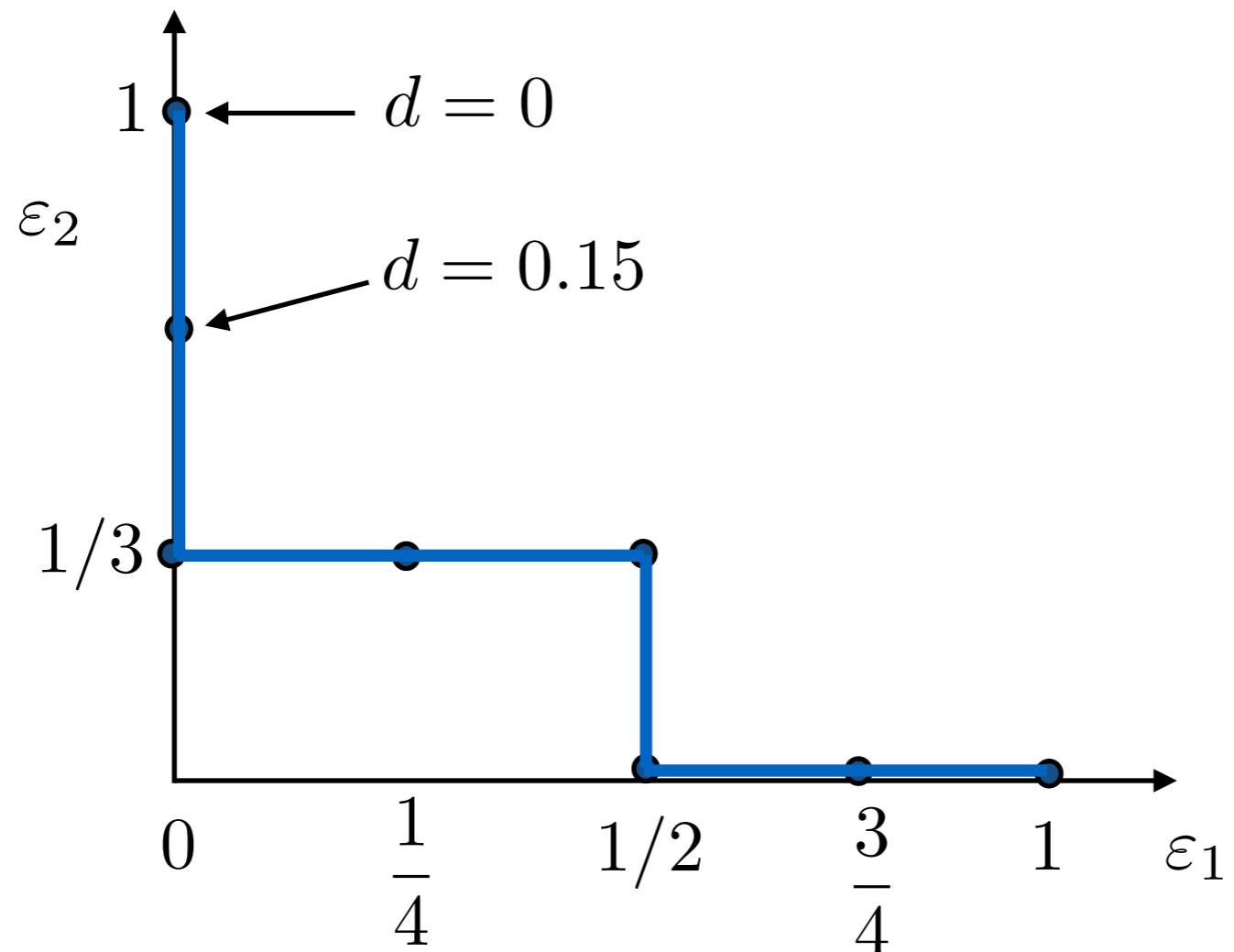
Example to compute ROC curve



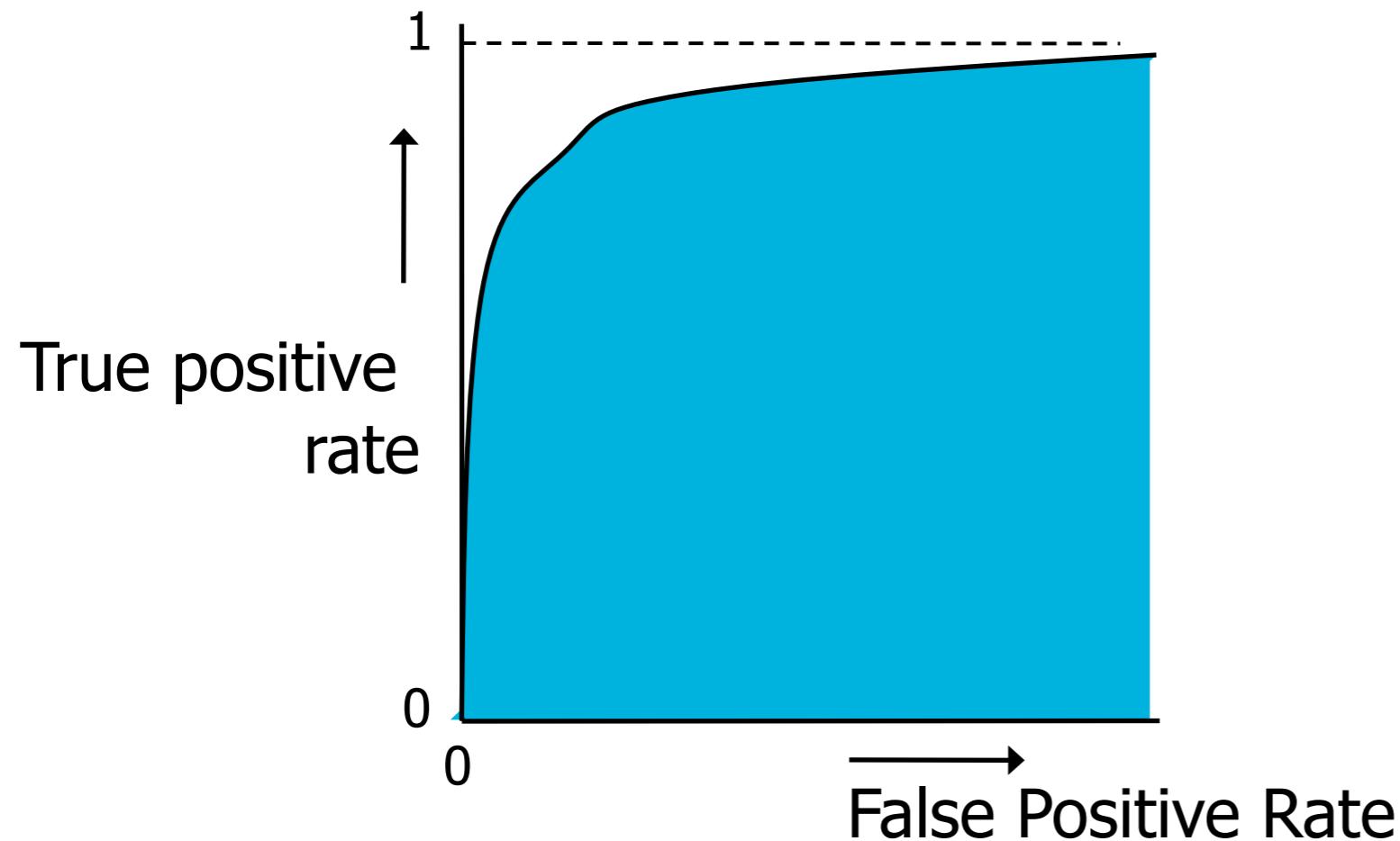
Example to compute ROC curve



Example to compute ROC curve



Area under the ROC curve: AUC



- Integrate the area: perfect classifier gives $AUC=1.0$
- Random classifier gives $AUC=0.5$
- Performance measure that is insensitive to class priors

Summary Reject and ROC

- Reject for solving ambiguity : reject objects close to the decision boundary → lower costs
- Reject option for coping with outliers
- ROC analysis in case of unknown or varying priors
- ROC analysis for comparing [and combining] classifiers

