

# Classification

# Today's outline

- Classification's difficulty
- Confusion matrix
- Precision, Recall
- F1 Score

# Classification's difficulty

What is the essential difference between regression and classification

# Classification

## Regression

- Predict a real value

| Feature |              | Target                    |
|---------|--------------|---------------------------|
| Income  | $h_{\theta}$ | House price               |
| 4.04    | $\mapsto$    | $h_{\theta}(4.04) = 2.02$ |
| 3.32    | $\mapsto$    | $h_{\theta}(3.32) = 1.66$ |
| 4.60    | $\mapsto$    | $h_{\theta}(4.60) = 2.30$ |
| 4.00    | $\mapsto$    | $h_{\theta}(4.00) = 2.00$ |

## Classification

- Predict a categorical value

| Feature |              | Target                                 |
|---------|--------------|--|
| Area    | $h_{\theta}$ | Tumour property                        |
| 13.64   | $\mapsto$    | $h_{\theta}(13.64) = \text{Benign}$    |
| 20.31   | $\mapsto$    | $h_{\theta}(20.31) = \text{Malignant}$ |
| 19.16   | $\mapsto$    | $h_{\theta}(19.16) = \text{Malignant}$ |
| 19.53   | $\mapsto$    | $h_{\theta}(19.53) = \text{Malignant}$ |

- Goal: find a reliable hypothesis  $h_{\theta}: (\text{feature}) \mapsto (\text{target})$

# Outline of classification

- Basically the same as that of regression
  - Design a model
  - Design a objective function
  - Optimisation
  - Evaluation
- Caution: naïve evaluation may lead us to a fatal failure...

# Naïve criterion: accuracy

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |

- $\text{accuracy} := \frac{\text{\# correct predictions}}{\text{\# total cases}}$ 
  - Ratio of correct predictions
- In the left example: 0.7
  - # correct predictions = 7
  - # total cases = 10
  - $\text{accuracy} = \frac{7}{10} = 0.7$
- Not always a good criterion

# “Cheating” to increase the accuracy

## Practical prediction

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

## Always say “benign”

| Predicted | Actual           | Performance |
|-----------|------------------|-------------|
| Benign    | Benign           | Correct     |
| Benign    | <b>Malignant</b> | Wrong       |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | <b>Malignant</b> | Wrong       |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | <b>Malignant</b> | Wrong       |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | <b>Malignant</b> | Wrong       |

Same accuracy

Beyond accuracy:  
Confusion matrix



# Two types of wrong predictions: Confusion matrix

- Accuracy does not distinguish diagnosing...
  - Malignant as benign
  - Benign as malignant
- Confusion matrix
  - How to evaluate? - precision, recall

|        |           | Predicted              |                        |
|--------|-----------|------------------------|------------------------|
|        |           | Malignant              | Benign                 |
| Actual | Malignant | True positive<br>(TP)  | False negative<br>(FN) |
|        | Benign    | False positive<br>(FP) | True negative<br>(TN)  |

# Confusion matrix: example

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

|        | Predicted |        |    |
|--------|-----------|--------|----|
|        | Malignant | Benign |    |
| Actual | Malignant | 3      | 1  |
|        | Benign    | 3      | 18 |

# Confusion matrix: example

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |

|        |           | Predicted |        |
|--------|-----------|-----------|--------|
|        |           | Malignant | Benign |
| Actual | Malignant | 3         | 1      |
|        | Benign    | 3         | 18     |

# Confusion matrix: example

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |

|        | Predicted |        |
|--------|-----------|--------|
|        | Malignant | Benign |
| Actual | Malignant | 3      |
|        | Benign    | 18     |

# Confusion matrix: example

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |

|        |           | Predicted |        |
|--------|-----------|-----------|--------|
|        |           | Malignant | Benign |
| Actual | Malignant | 3         | 1      |
|        | Benign    | 3         | 18     |

# Confusion matrix: example

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |

|        | Predicted |        |    |
|--------|-----------|--------|----|
|        | Malignant | Benign |    |
| Actual | Malignant | 3      | 1  |
|        | Benign    | 3      | 18 |

# Confusion matrix: example

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

|        |           | Predicted |        |
|--------|-----------|-----------|--------|
|        |           | Malignant | Benign |
| Actual | Malignant | 3         | 1      |
|        | Benign    | 3         | 18     |

How can we define a score on this matrix?  
- Precision, Recall

# Precision and recall

Let's quantify the confusion matrix!



# Precision:

|        |           | Predicted                   |                             |                            |
|--------|-----------|-----------------------------|-----------------------------|----------------------------|
|        |           | Malignant                   | Benign                      |                            |
| Actual | Malignant | True positive (TP)          | False negative (FN)         | Actually positive<br>TP+FN |
|        | Benign    | False positive (FP)         | True negative (TN)          | Actually negative<br>FP+TN |
| Total  |           | Predicted positive<br>TP+FP | Predicted negative<br>FN+TN | Total cases<br>TP+FP+FN+TN |

| Precision | $TP/(TP+FP)$ |
|-----------|--------------|
|-----------|--------------|

| Recall       |
|--------------|
| $TP/(TP+FN)$ |

| F1 score              |
|-----------------------|
| $2/(1/pre. + 1/rec.)$ |

# Precision: example

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

# Precision: example

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |

|           | Predicted |        | Total |
|-----------|-----------|--------|-------|
|           | Malignant | Benign |       |
| Malignant | 3         | 1      | 4     |
| Benign    | 3         | 18     | 21    |
| Total     | 6         | 19     | 25    |

| Precision | 0.50 |
|-----------|------|
|-----------|------|

| Recall |
|--------|
| 0.75   |

| F1 score |
|----------|
| 0.60     |

# Recall:

|        |           | Predicted                   |                             | Total                      |
|--------|-----------|-----------------------------|-----------------------------|----------------------------|
|        |           | Malignant                   | Benign                      |                            |
| Actual | Malignant | True positive (TP)          | False negative (FN)         | Actually positive<br>TP+FN |
|        | Benign    | False positive (FP)         | True negative (TN)          | Actually negative<br>FP+TN |
| Total  |           | Predicted positive<br>TP+FP | Predicted negative<br>FN+TN | Total cases<br>TP+FP+FN+TN |

|           |              |
|-----------|--------------|
| Precision | $TP/(TP+FP)$ |
|-----------|--------------|

|              |
|--------------|
| Recall       |
| $TP/(TP+FN)$ |

|                       |
|-----------------------|
| F1 score              |
| $2/(1/pre. + 1/rec.)$ |

# Recall: example

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

# Recall: example

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Benign    | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |

|        |           | Predicted |        | Total |
|--------|-----------|-----------|--------|-------|
|        |           | Malignant | Benign |       |
| Actual | Malignant | 3         | 1      | 4     |
|        | Benign    | 3         | 18     | 21    |
| Total  |           | 6         | 19     | 25    |

|           |      |
|-----------|------|
| Precision | 0.50 |
|-----------|------|

|        |
|--------|
| Recall |
| 0.75   |

|          |
|----------|
| F1 score |
| 0.60     |

# Accuracy vs precision and recall

| Correct | Wrong | Total       | Accuracy               |
|---------|-------|-------------|------------------------|
| TP+TN   | FP+FN | TP+FP+FN+TN | $(TP+TN)/\text{Total}$ |

|        | Predicted                   |                             | Total                      | Recall       |
|--------|-----------------------------|-----------------------------|----------------------------|--------------|
|        | Malignant                   | Benign                      |                            |              |
| Actual | True positive (TP)          | False negative (FN)         | Actually positive<br>TP+FN | $TP/(TP+FN)$ |
|        | False positive (FP)         | True negative (TN)          | Actually negative<br>FP+TN |              |
| Total  | Predicted positive<br>TP+FP | Predicted negative<br>FN+TN | Total cases<br>TP+FP+FN+TN |              |

| Precision    | F1 score                            |
|--------------|-------------------------------------|
| $TP/(TP+FP)$ | $2/(1/\text{pre.} + 1/\text{rec.})$ |

# If we always say “benign”

| Predicted | Actual           | Performance |
|-----------|------------------|-------------|
| Benign    | Benign           | Correct     |
| Benign    | <b>Malignant</b> | Wrong       |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | <b>Malignant</b> | Wrong       |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | <b>Malignant</b> | Wrong       |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | Benign           | Correct     |
| Benign    | <b>Malignant</b> | Wrong       |



If we always say “benign”

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |

|        | Predicted |        | Total |
|--------|-----------|--------|-------|
|        | Malignant | Benign |       |
| Actual | Malignant | 0      | 4     |
|        | Benign    | 0      | 21    |
| Total  |           | 0      | 25    |

Precision #DIV/0!

# If we always say “benign”

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |

|        |           | Predicted |        | Total |
|--------|-----------|-----------|--------|-------|
|        |           | Malignant | Benign |       |
| Actual | Malignant | 0         | 4      | 4     |
|        | Benign    | 0         | 21     | 21    |
| Total  |           | 0         | 25     | 25    |

|           |         |
|-----------|---------|
| Precision | #DIV/0! |
|-----------|---------|

|        |
|--------|
| Recall |
| 0.00   |

|          |
|----------|
| F1 score |
| #DIV/0!  |

# If we detect one malignant...

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

If we detect one malignant...

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |

|        | Predicted |        | Total |    |
|--------|-----------|--------|-------|----|
|        | Malignant | Benign |       |    |
| Actual | Malignant | 1      | 3     | 4  |
|        | Benign    | 0      | 21    | 21 |
| Total  |           | 1      | 24    | 25 |

|           |      |
|-----------|------|
| Precision | 1.00 |
|-----------|------|

# If we detect one malignant...

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Malignant | Wrong       |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Benign    | Benign    | Correct     |
| Malignant | Malignant | Correct     |

|        |           | Predicted |        | Total |
|--------|-----------|-----------|--------|-------|
|        |           | Malignant | Benign |       |
| Actual | Malignant | 1         | 3      | 4     |
|        | Benign    | 0         | 21     | 21    |
| Total  |           | 1         | 24     | 25    |

|           |      |
|-----------|------|
| Precision | 1.00 |
|-----------|------|

|        |
|--------|
| Recall |
| 0.25   |

|          |
|----------|
| F1 score |
| 0.40     |

# If we always say “malignant”

| Predicted | Actual    | Performance |
|-----------|-----------|-------------|
| Malignant | Benign    | Wrong       |
| Malignant | Malignant | Correct     |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Malignant | Correct     |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Malignant | Correct     |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Benign    | Wrong       |
| Malignant | Malignant | Correct     |

|        | Predicted |        | Total |
|--------|-----------|--------|-------|
|        | Malignant | Benign |       |
| Actual | Malignant | 4      | 4     |
|        | Benign    | 21     | 21    |
| Total  |           | 25     | 25    |

| Precision | 0.16 |
|-----------|------|
|-----------|------|

| Recall |
|--------|
| 1.00   |

| F1 score |
|----------|
| 0.28     |

# If we always say “malignant”

| Predicted        | Actual           | Performance    |
|------------------|------------------|----------------|
| Malignant        | Benign           | Wrong          |
| <b>Malignant</b> | <b>Malignant</b> | <b>Correct</b> |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| <b>Malignant</b> | <b>Malignant</b> | <b>Correct</b> |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| <b>Malignant</b> | <b>Malignant</b> | <b>Correct</b> |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| Malignant        | Benign           | Wrong          |
| <b>Malignant</b> | <b>Malignant</b> | <b>Correct</b> |

|        |           | Predicted |        | Total |
|--------|-----------|-----------|--------|-------|
|        |           | Malignant | Benign |       |
| Actual | Malignant | 4         | 0      | 4     |
|        | Benign    | 21        | 0      | 21    |
| Total  |           | 25        | 0      | 25    |

|           |      |
|-----------|------|
| Precision | 0.16 |
|-----------|------|

|        |
|--------|
| Recall |
| 1.00   |

|          |
|----------|
| F1 score |
| 0.28     |

Integrate precision and Recall:  
F1 score



# How to compare?

- The naïve mean of the precision and recall is not suitable.

Practical prediction

Precision: 0.500

Recall: 0.750

Mean: 0.625

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

Negative-based

Precision: 1.000

Recall: 0.250

Mean: 0.625

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

# F1 score

- Extremely low precision or recall is impractical.
- Idea: penalising extremely low values strongly.
  - Take the reciprocal of the mean of the reciprocals:

$$(\text{F1 score}) := \frac{1}{\frac{1}{2} \left( \frac{1}{(\text{precision})} + \frac{1}{(\text{recall})} \right)}$$

- Example: if (precision) = 1.0, (recall) = 0.01 (cf. (Mean) = 0.505)

$$(\text{F1 score}) = \frac{1}{\frac{1}{2} \left( \frac{1}{1.0} + \frac{1}{0.01} \right)} = \frac{1}{\frac{1}{2} (1 + 100)} = \frac{1}{50.5} \approx 0.02$$

Using reciprocals penalises extremely low values

# F1 score: examples

- F1 score penalises extremely low precision and recall

Practical prediction

Precision: 0.500

Recall: 0.750

Mean: 0.625

F1 score: 0.60

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | Benign           | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |

Negative-based

Precision: 1.000

Recall: 0.250

Mean: 0.625

F1 score: 0.40

| Predicted        | Actual           | Performance |
|------------------|------------------|-------------|
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | <b>Malignant</b> | Wrong       |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| Benign           | Benign           | Correct     |
| <b>Malignant</b> | <b>Malignant</b> | Correct     |