Outline

Confusion or error matrix

- Binary scenario $\mathcal{W} = \{\omega_1, \omega_2\} \equiv \{\omega, \overline{\omega}\}$

Confusion or error matrix

- Binary scenario $\mathcal{W} = \{\omega_1, \omega_2\} \equiv \{\omega, \overline{\omega}\}$
- Confusion / error matrix:

Expected	Prediction ω	Prediction $\overline{\omega}$
ω	True Positive (TP)	False Negative (FN)
$\overline{\omega}$	False Positive (FP)	True Negative (TN)

Confusion or error matrix

- Binary scenario $\mathcal{W} = \{\omega_1, \omega_2\} \equiv \{\omega, \overline{\omega}\}$
- Confusion / error matrix:

Expected	Prediction ω	Prediction $\overline{\omega}$
ω	True Positive (TP)	False Negative (FN)
$\overline{\omega}$	False Positive (FP)	True Negative (TN)

- Trained classification model $\hat{f} : \mathbb{R}^d \to \mathcal{W}$
- Test dataset $\Rightarrow \mathcal{D}_{test} = \mathcal{D}_{\omega} \cup \mathcal{D}_{\overline{\omega}}$

Confusion or error matrix

- Binary scenario $\mathcal{W} = \{\omega_1, \omega_2\} \equiv \{\omega, \overline{\omega}\}$
- Confusion / error matrix:

Expected	Prediction ω	Prediction $\overline{\omega}$
ω	True Positive (TP)	False Negative (FN)
$\overline{\omega}$	False Positive (FP)	True Negative (TN)

- Trained classification model $\hat{f} : \mathbb{R}^d \rightarrow \mathcal{W}$
- Test dataset $\Rightarrow \mathcal{D}_{test} = \mathcal{D}_\omega \cup \mathcal{D}_{\overline{\omega}}$

TP: $\sum_{\mathbf{x} \in \mathcal{D}_\omega} [\![\hat{f}(\mathbf{x}) = \omega]\!]$ FN: $\sum_{\mathbf{x} \in \mathcal{D}_\omega} [\![\hat{f}(\mathbf{x}) = \overline{\omega}]\!]$

TN: $\sum_{\mathbf{x} \in \mathcal{D}_{\overline{\omega}}} [\![\hat{f}(\mathbf{x}) = \overline{\omega}]\!]$ FP: $\sum_{\mathbf{x} \in \mathcal{D}_{\overline{\omega}}} [\![\hat{f}(\mathbf{x}) = \omega]\!]$

Confusion or error matrix

- Binary scenario $\mathcal{W} = \{\omega_1, \omega_2\} \equiv \{\omega, \overline{\omega}\}$
- Confusion / error matrix:

Expected	**Prediction** ω	$\overline{\omega}$
ω	True Positive (TP)	False Negative (FN)
$\overline{\omega}$	False Positive (FP)	True Negative (TN)

- Trained classification model $\hat{f} : \mathbb{R}^d \to \mathcal{W}$
- Test dataset $\Rightarrow \mathcal{D}_{test} = \mathcal{D}_\omega \cup \mathcal{D}_{\overline{\omega}}$

TP: $\sum_{\mathbf{x} \in \mathcal{D}_\omega} [\![\hat{f}(\mathbf{x}) = \omega]\!]$ FN: $\sum_{\mathbf{x} \in \mathcal{D}_\omega} [\![\hat{f}(\mathbf{x}) = \overline{\omega}]\!]$ $|\mathcal{D}_\omega| = \text{TP} + \text{FN}$

TN: $\sum_{\mathbf{x} \in \mathcal{D}_{\overline{\omega}}} [\![\hat{f}(\mathbf{x}) = \overline{\omega}]\!]$ FP: $\sum_{\mathbf{x} \in \mathcal{D}_{\overline{\omega}}} [\![\hat{f}(\mathbf{x}) = \omega]\!]$ $|\mathcal{D}_{\overline{\omega}}| = \text{TN} + \text{FP}$

Examples of confusion matrices

Engineer #1

Expected	**Prediction** Sunny	Rainy
Sunny	350	5
Rainy	5	5

Engineer #2

Expected	**Prediction** Sunny	Rainy
Sunny	355	0
Rainy	10	0

Metrics

Metrics

1. Accuracy (Acc)
   - Ratio between correct predictions and total number of guesses:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} = \frac{\text{TP} + \text{TN}}{|\mathcal{D}_{\omega}| + |\mathcal{D}_{\overline{\omega}}|}$$

   - Errors are equally weighted
   - Suitable for balanced scenarios

Metrics

1. Accuracy (Acc)
   - Ratio between correct predictions and total number of guesses:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} = \frac{\text{TP} + \text{TN}}{|\mathcal{D}_\omega| + |\mathcal{D}_{\overline{\omega}}|}$$

   - Errors are equally weighted
   - Suitable for balanced scenarios

2. Precision (P)
   - Correct claims VS all positive predictions:

$$\text{P} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

   - Penalizes false alarms

3. Recall (R)
   - Correct claims VS expected positive predictions:

$$\text{R} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

   - Penalizes missed positives

Metrics

4. F-measure (F_β)
   - P and R optimize different aspects $\Rightarrow$ need for a single indicator
   - Harmonic mean between P and R:

$$F_\beta = (1+\beta^2) \cdot \frac{P \cdot R}{\beta^2 \cdot P + R}$$

   - Most commonly, $\beta = 1$:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

Metrics

4. F-measure (F_β)
   - P and R optimize different aspects $\Rightarrow$ need for a single indicator
   - Harmonic mean between P and R:

$$F_\beta = (1+\beta^2)\cdot\frac{P\cdot R}{\beta^2\cdot P + R}$$

   - Most commonly, $\beta = 1$:

$$F_1 = 2\cdot\frac{P\cdot R}{P+R} = \frac{2\cdot TP}{2\cdot TP + FP + FN}$$

$\Rightarrow$ Considerations:

- Acc: Global for all classes
- P, R, F_1: Computed for each individual class

Exercise

Engineer #1

Expected	**Prediction** Sunny	Rainy
Sunny	350	5
Rainy	5	5

Engineer #2

Expected	**Prediction** Sunny	Rainy
Sunny	355	0
Rainy	10	0

Metrics (ii)

5. True Positive Rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

- Also: *Recall*, *Sensitivity*

6. True Negative Rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

- Also: *Specificity*

Metrics (ii)

5. True Positive Rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

- Also: *Recall*, *Sensitivity*

6. True Negative Rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

- Also: *Specificity*

7. False Positive Rate (FPR)

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

8. False Negative Rate (FNR)

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TP}}$$

Metrics (ii)

5. True Positive Rate (TPR)

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- Also: *Recall*, *Sensitivity*

6. True Negative Rate (TNR)

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

- Also: *Specificity*

7. False Positive Rate (FPR)

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

8. False Negative Rate (FNR)

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}}$$

$$\text{TPR} + \text{FNR} = 1$$
$$\text{TNR} + \text{FPR} = 1$$

Summary

Figure of merit	Formula	Computation
Accuracy (Acc)	$^{(TP+TN)}/_{(TP+FN+TN+FP)}$	Global
Precision (P)	$^{TP}/_{(TP+FP)}$	Class-wise
Recall (R)	$^{TP}/_{(TP+FN)}$	Class-wise
F-measure (F_1)	$^{2 \cdot TP}/_{(2 \cdot TP+FP+FN)}$	Class-wise
True Positive Rate (TPR)	$^{TP}/_{(TP+FN)}$	Class-wise
True Negative Rate (TNR)	$^{TN}/_{(TN+FP)}$	Class-wise
False Positive Rate (FPR)	$^{FP}/_{(FP+TN)}$	Class-wise
False Negative Rate (FNR)	$^{FN}/_{(FN+TP)}$	Class-wise

Error trade-offs

- Decision thresholds remarkably impacts the recognition performance

Error trade-offs

- Decision thresholds remarkably impacts the recognition performance

- Intuitively:
  - $\rightarrow$ **Lower** threshold: Increase in Positive predictions (TPR $\uparrow$, FPR $\uparrow$)
  - $\rightarrow$ **Higher** thresholds: Decrease in Positive predictions (TPR $\downarrow$, FPR $\downarrow$)

Error trade-offs

- Decision thresholds remarkably impacts the recognition performance

- Intuitively:
  - $\rightarrow$ **Lower** threshold: Increase in Positive predictions (TPR $\uparrow$, FPR $\uparrow$)
  - $\rightarrow$ **Higher** thresholds: Decrease in Positive predictions (TPR $\downarrow$, FPR $\downarrow$)

- Each threshold corresponds to a particular duple (TPR, FPR)

Receiver Operating Characteristic and Area Under Curve

ROC curve: FPR against TPR as the decision threshold varies

- Top-left corner (FPR=0, TPR=1): perfect model
- Diagonal line (FPR = TPR): random guessing
- Below diagonal: worse than random

Receiver Operating Characteristic and Area Under Curve

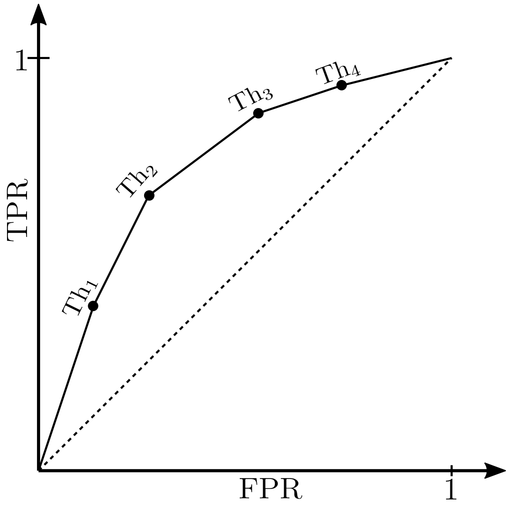
ROC curve: FPR against TPR as the decision threshold varies

- Top-left corner (FPR=0, TPR=1): perfect model
- Diagonal line (FPR = TPR): random guessing
- Below diagonal: worse than random

AUC: Area under the ROC curve

- Independent of class imbalance
- Threshold-free metric
- Ranges:
  - → $0.5 < \text{AUC} \leq 1 \Rightarrow$ Performs adequately
  - → $\text{AUC} = 0.5 \Rightarrow$ Random guessing
  - → $0 \leq \text{AUC} < 0.5 \Rightarrow$ Underperforming

Receiver Operating Characteristic and Area Under Curve



Introduction

- In binary scenarios ($|\mathcal{W}| = 2$), one class is selected as positive
  - $\rightarrow$ Multiclass ($|\mathcal{W}| > 2$): each class can be positive $\Rightarrow$ one-vs-all

- Accuracy requires no adaptation

Introduction

- In binary scenarios ($|\mathcal{W}| = 2$), one class is selected as positive
  - $\rightarrow$ Multiclass ($|\mathcal{W}| > 2$): each class can be positive $\Rightarrow$ one-vs-all
- Accuracy requires no adaptation
- Averaging process to summarize the performance across all classes
  - $\rightarrow$ Different averaging strategies

Adaptations

1. Micro-Average

- Aggregates at the error level:

$$\text{Micro-P} = \frac{\sum_i \text{FP}_i}{\sum_i (\text{TP}_i + \text{FP}_i)} \quad \text{Micro-R} = \frac{\sum_i \text{FP}_i}{\sum_i (\text{TP}_i + \text{FN}_i)} \quad \text{Micro-F}_1 = \frac{2 \cdot \sum_i \text{TP}_i}{\sum_i (2\text{TP}_i + \text{FP}_i + \text{FN}_i)}$$

- Treats individual predictions equally
- Favors majority classes

Adaptations

1. Micro-Average
   - Aggregates at the error level:

$$\text{Micro-P} = \frac{\sum_i \text{FP}_i}{\sum_i (\text{TP}_i + \text{FP}_i)} \quad \text{Micro-R} = \frac{\sum_i \text{FP}_i}{\sum_i (\text{TP}_i + \text{FN}_i)} \quad \text{Micro-F}_1 = \frac{2 \cdot \sum_i \text{TP}_i}{\sum_i (2\text{TP}_i + \text{FP}_i + \text{FN}_i)}$$

   - Treats individual predictions equally
   - Favors majority classes

2. Macro-Average (weighted)
   - Aggregates at the metric level:

$$\text{Macro-P} = \frac{1}{|\mathcal{W}|} \sum_{i=1}^{|\mathcal{W}|} \epsilon_i \cdot \text{P}_i \qquad \text{Macro-R} = \frac{1}{|\mathcal{W}|} \sum_{i=1}^{|\mathcal{W}|} \epsilon_i \cdot \text{R}_i \qquad \text{Macro-F}_1 = \frac{1}{|\mathcal{W}|} \sum_{i=1}^{|\mathcal{W}|} \epsilon_i \cdot \text{F}_{1i}$$

   - General case: $\epsilon_i = {}^{|\mathcal{D}_i|}/_{|\mathcal{D}|} \Rightarrow$ Without weighting: $\epsilon_i = 1$
   - Treats all classes equally ($\epsilon_i = 1$)
   - Highlights performance on minority classes ($\epsilon_i = 1$)

Exercise

Expected	Prediction: Sunny	Rainy	Windy	Cloudy
Sunny	285	5	5	10
Rainy	2	10	2	1
Windy	0	5	15	10
Cloudy	0	3	2	10

Extensions to other scenarios

- **Ordinal classification**: Typically evaluated as a regression task

Extensions to other scenarios

- **Ordinal classification**: Typically evaluated as a regression task

- **Multilabel classification**:
  - → Adapted metrics: Macro-averaging (average per label), micro-averaging (aggregate over labels)
  - → *Ad-hoc* metrics: Hamming Loss, Jaccard Index

Outline

Introduction

- Continuous target $\Rightarrow \mathcal{W} \subseteq \mathbb{R}$

Introduction

- Continuous target $\Rightarrow \mathcal{W} \subseteq \mathbb{R}$
  - $\rightarrow$ How far the prediction deviates from the expected value

Introduction

- Continuous target $\Rightarrow \mathcal{W} \subseteq \mathbb{R}$
  - $\rightarrow$ How far the prediction deviates from the expected value

- Trained model $\hat{f} : \mathbb{R}^d \rightarrow \mathcal{W}$
- Test dataset $\Rightarrow \mathcal{D}_{test} = \{(\mathbf{x}_i, \omega_i)\}_{i=1}^{|\mathcal{D}_{test}|}$
  - $\rightarrow$ Feature vector $\mathbf{x} \in \mathbb{R}^d$

Main metrics

1. Mean Absolute Error (MAE)
   - Aggregates at the error level:

$$\mathrm{MAE} = \frac{1}{\left|\mathcal{D}_{test}\right|} \sum_{i=1}^{\left|\mathcal{D}_{test}\right|} \left|\hat{f}(\mathbf{x}_i) - \omega_i\right|$$

Main metrics

1. Mean Absolute Error (MAE)
   - Aggregates at the error level:

$$\text{MAE} = \frac{1}{|\mathcal{D}_{test}|} \sum_{i=1}^{|\mathcal{D}_{test}|} \left| \hat{f}(\mathbf{x}_i) - \omega_i \right|$$

2. Mean Squared Error (MSE)
   - Aggregates at the metric level:

$$\text{MSE} = \frac{1}{|\mathcal{D}_{test}|} \sum_{i=1}^{|\mathcal{D}_{test}|} \left(\hat{f}(\mathbf{x}_i) - \omega_i \right)^2$$

T3: Model evaluation

Fundamentos del Aprendizaje Automático

Curso 2025/2026