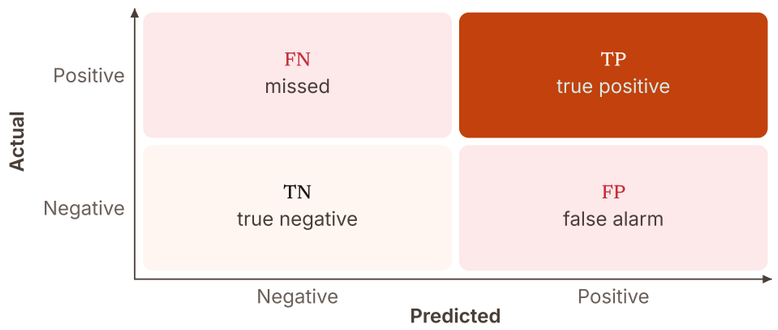
Logistic regression, overfitting and validation

POSTER · THE MODULE AT A GLANCE

Exam Mon, 7 Dec 2026 · 180 min

1 Confusion matrix



Accuracy = (TP + TN) / n. Sensitivity = TP / (TP + FN). Specificity = TN / (TN + FP). Precision = TP / (TP + FP).

2 Preventing overfitting with validation

1 Split into train and test → 2 Fit on the train set → 3 Predict the test set → 4 Compare train and test performance

Judge a model on data it has not seen.

EXAM LIKELIHOOD

Logistic regression (logit model)	98%
Confusion matrix, computing accuracy and sensitivity	97%
Accuracy is misleading with imbalanced classes	95%
A logit coefficient is a change in log-odds, not in probability	94%
Odds and odds ratio (e^b)	93%
Overfitting and generalization	92%

COMMON MISTAKES

- ✖ p = 0.06 is not a 6% chance of H0
- ✖ "2.75 times the chance" is wrong: it is odds
- ✖ 0.5 is a habit, not a rule
- ✖ A p-value says nothing about the size of the effect
- ✖ MLE properties hold for large n

HOW TO WRITE THE ANSWER

1. Term — model or concept, English term

2. Definition — definition or features, precisely

3. Case fact — concrete fact from the case

4. Conclusion — judgement, advice or consequence

KEY CONCEPTS

logistic regression · binary dependent variable · log-odds · odds ratio · maximum likelihood · AIC · McFadden pseudo-R^2 · confusion matrix · accuracy · sensitivity · specificity · precision · threshold · ROC-curve · AUC · overfitting · cross-validation · bias-variance trade-off