Lecture 10: Regression Metrics

Firas Moosvi (Slides adapted from Varada Kolhatkar)

Which metric fits best?

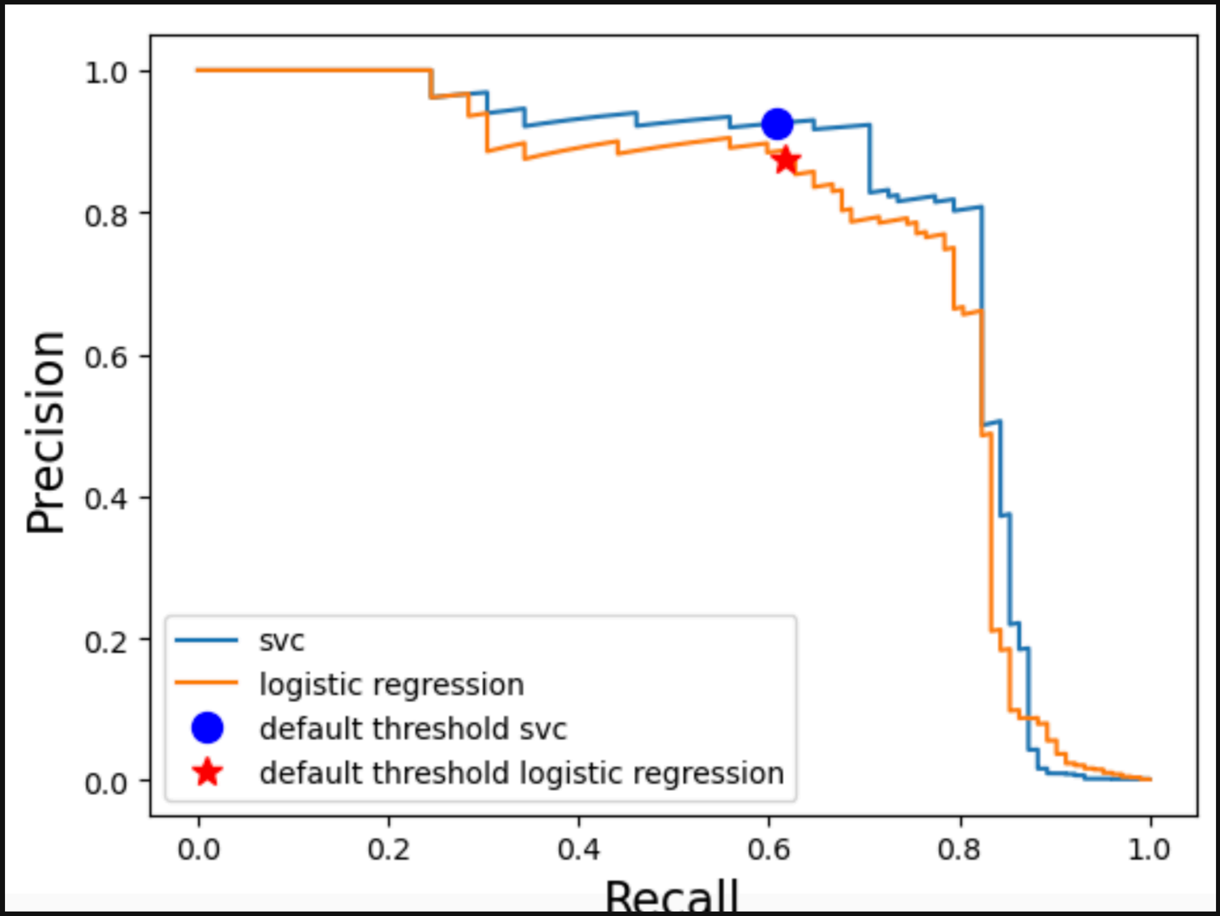
Scenario	Data Imbalance	Main Concern	Best Metric(s) / Curve
Email Spam Detection	10% spam	Avoid false positives	
Disease Screening	1 in 10,000	Avoid false negatives	
Credit Card Fraud	0.1% fraud	Focus on rare positive class	
Customer Churn	20% churn	Balance FP & FN	
Sentiment Analysis	50/50 balanced	Overall correctness	
Face Recognition	Balanced pairs	Trade-off FP vs FN	

Summary: Choosing the right metric

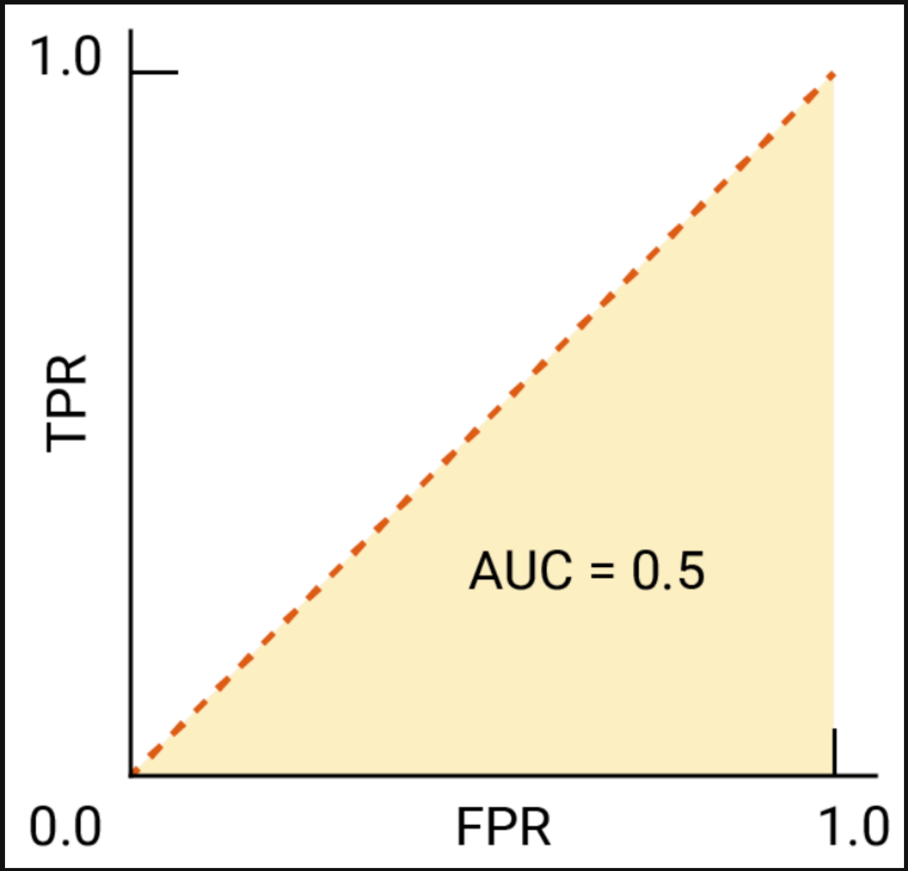
Metric / Plot	When to Use	Why
Precision, Recall, F1	When you care about *specific error types* (FP vs FN) or a *fixed threshold*.	Focus on particular tradeoffs.
PR Curve & AP Score	When the dataset is **highly imbalanced** (rare positives).	Ignores TNs; focuses on positives.
ROC Curve & AUC	When classes are **moderately imbalanced**.	Measures ranking ability across thresholds.

Questions for you

- What's the difference between the average precision (AP) score and F1-score?
- Which model would you pick?



ROC of a baseline model

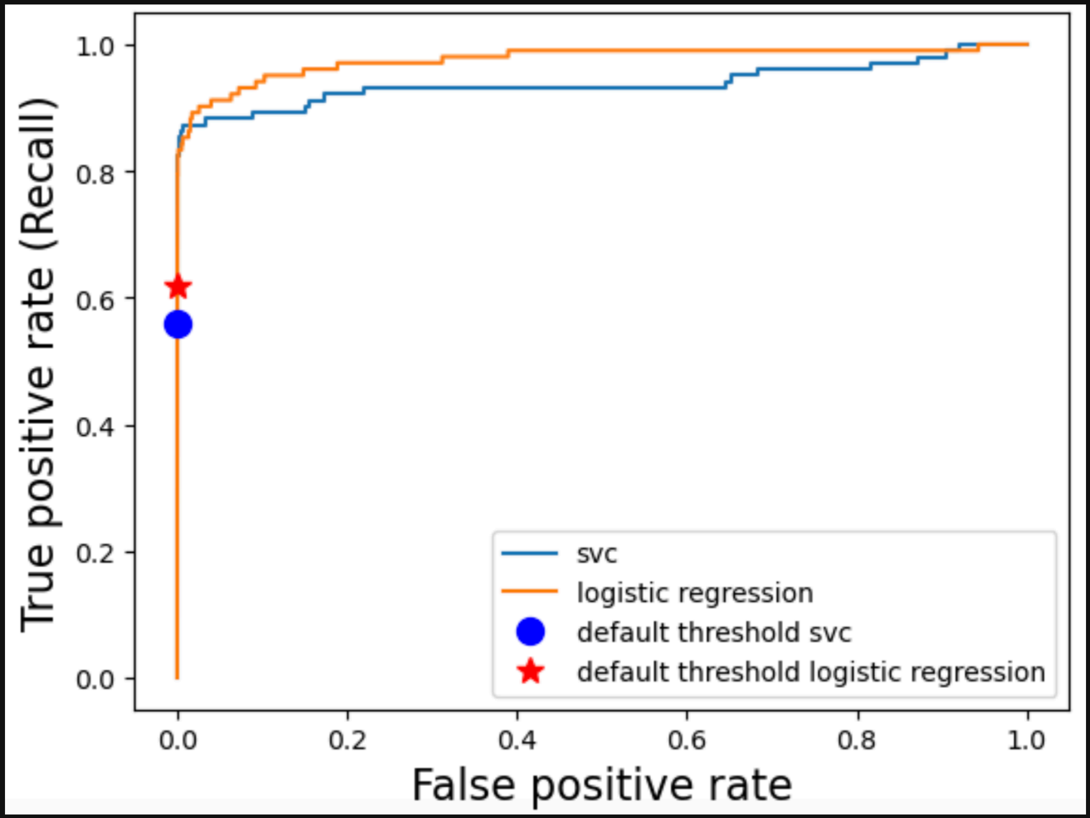


AUC–ROC measures the probability that a randomly chosen **positive example** receives a **higher score** than a randomly chosen **negative example**.

- **Perfect model:** (AUC = 1.0). Always ranks positives above negatives.
- **Random model** (AUC = 0.5): No discriminative ability (equivalent to random guessing).

Source

Questions for you



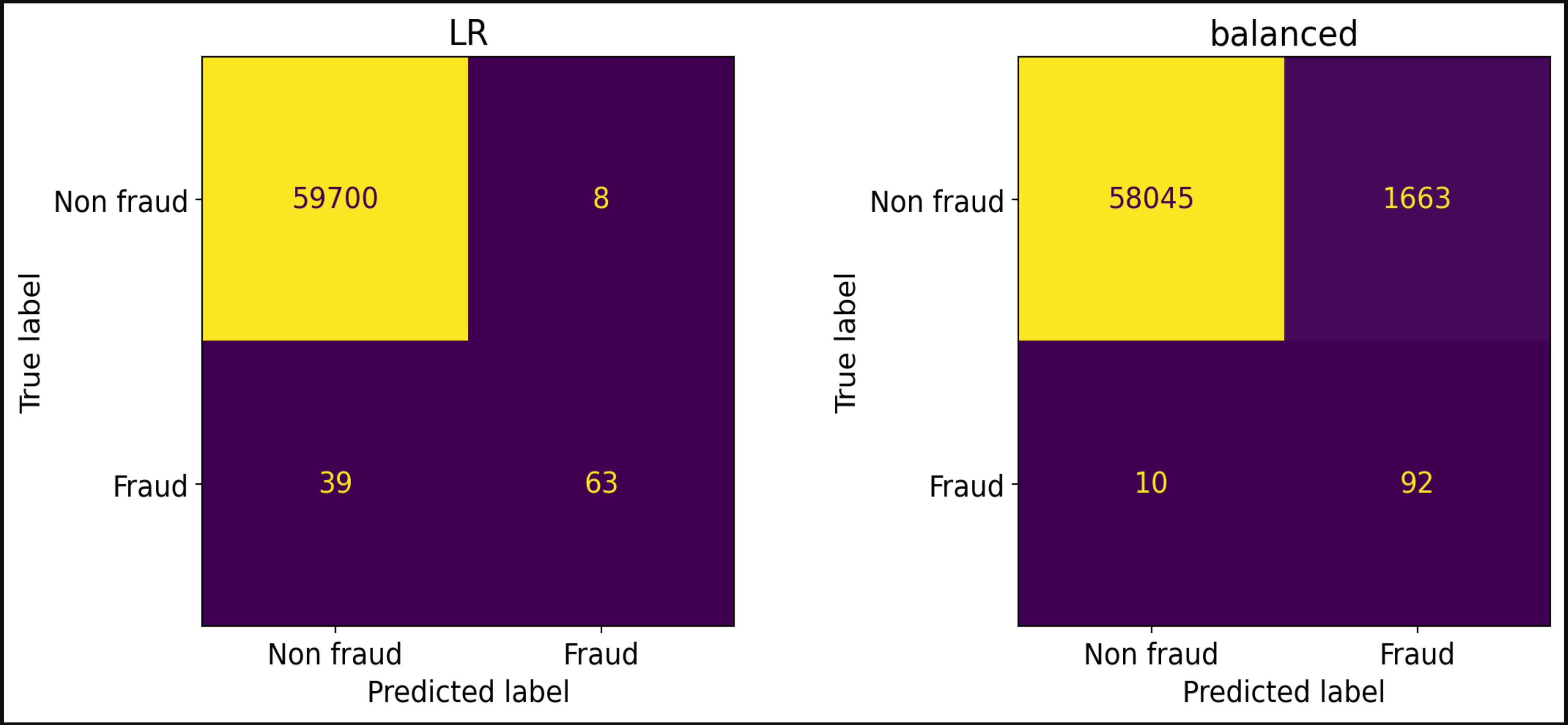
- Which model would you pick?

Dealing with class imbalance

- Under sampling
- Oversampling
- `class weight="balanced"` (preferred method for this course)
- SMOTE

Handling imbalance by chaning class weights

- We can specify class_weight="balanced" to give more importance to rare examples during training.



Ridge and RidgeCV

- **Ridge Regression**: `alpha` hyperparameter controls model complexity.
- **RidgeCV**: Ridge regression with built-in cross-validation to find the optimal `alpha`.

`alpha` hyperparameter

- **Role of `alpha`**:
  - Controls model complexity
  - Higher `alpha`: Simpler model, smaller coefficients.
  - Lower `alpha`: Complex model, larger coefficients.

Regression metrics: MSE, RMSE, MAPE, r2_score

- **Mean Squared Error (MSE)**: Average of the squares of the errors.
- **Root Mean Squared Error (RMSE)**: Square root of MSE, same units as the target variable.
- **r2** measures how much of the variation in the target variable your model can explain.-
- **Mean Absolute Percentage Error (MAPE)**: Average of the absolute percentage errors.

Applying log transformation to the targets

- Suitable when the target has a wide range and spans several orders of magnitude
  - Example: counts data such as social media likes or price data
- Helps manage skewed data, making patterns more apparent and regression models more effective.
- `TransformedTargetRegressor`
  - Wraps a regression model and applies a transformation to the target values.

iClicker Exercise 10.1

Select all of the following statements which are TRUE.

- a. Price per square foot would be a good feature to add in our `X`.
- b. The `alpha` hyperparameter of `Ridge` has similar interpretation of `C` hyperparameter of `LogisticRegression`; higher `alpha` means more complex model.
- c. In `Ridge`, smaller alpha means bigger coefficients whereas bigger alpha means smaller coefficients.

Group Work: Class Demo & Live Coding

For this demo, each student should click this link to create a new repo in their accounts, then clone that repo locally to follow along with the demo from today.

Which metric fits the scenario?

Scenario	What matters most?	Best metric(s)?
Predicting **house prices** ranging from $60K–$800K.	A $30K error is huge for a $60K house but small for a $500K house.	
Predicting **exam scores** (0–100).	You want an interpretable measure of average error in **points**.	
Predicting **energy consumption** in a large industrial system.	Large errors are very costly and should be penalized heavily.	
Predicting **insurance claim amounts**.	You want to compare how well different models explain the variation in claims.	

iClicker Exercise 10.2

Select all of the following statements which are TRUE.

- a. We can use still use precision and recall for regression problems but now we have other metrics we can use as well.
- b. In `sklearn` for regression problems, using `r2_score()` and `.score()` (with default values) will produce the same results.
- c. RMSE is always going to be non-negative.
- d. MSE does not directly provide the information about whether the model is underpredicting or overpredicting.
- e. We can pass multiple scoring metrics to `GridSearchCV` or `RandomizedSearchCV` for regression as well as classification problems.

Which metric fits the scenario?

Scenario	What matters most?	Best metric(s)?
Predicting **house prices** ranging from \$60K–\$800K.	A \$30K error is huge for a \$60K house but small for a \$500K house.	
Predicting **exam scores** (0–100).	You want an interpretable measure of average error in **points**.	
Predicting **energy consumption** in a large industrial system.	Large errors are very costly and should be penalized heavily.	
Predicting **insurance claim amounts**.	You want to compare how well different models explain the variation in claims.	

Which metric fits the scenario?

- **For interpretability:** prefer RMSE or MAPE
- **When you want to discourage large error:** MSE is common
- **For fair comparison:** r2 provides a normalized score similar to accuracy in classification.
- **For imbalanced scales:** MAPE helps when proportional error matters more than absolute error.