Workshop: Machine Learning Processing for Wearable Data in Healthcare

Classification and Regression Cases in Rehabilitation Event Detection

Dr Diego Paez
Mehdi Ejtehadi, Yanke Li, Bertram Fuchs

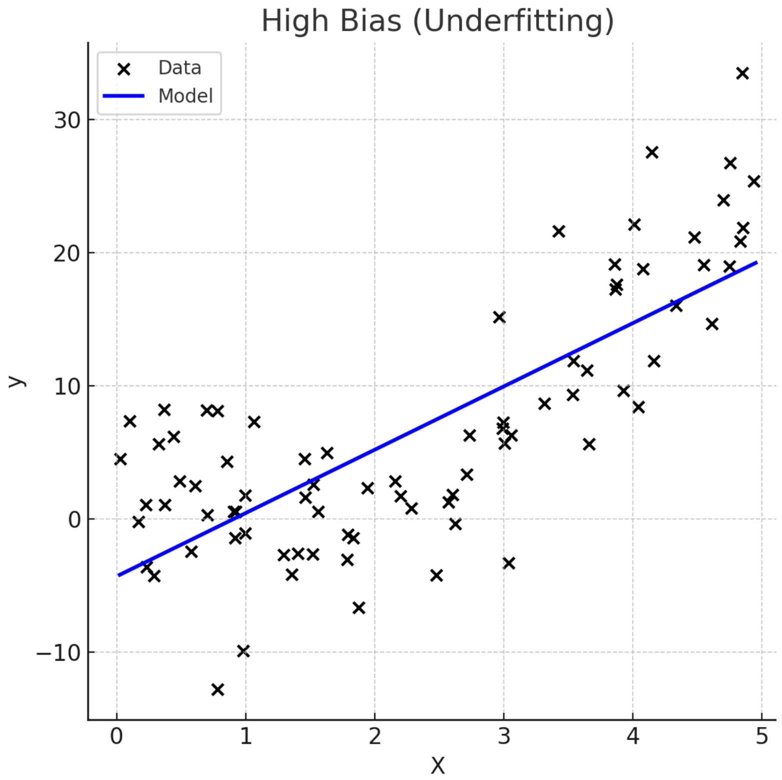
WS Day 2: Evaluation Metrics for Acceptable Machine Learning

Content and Learning Outcomes:

1. Data-driven model principles: Bias variance trade-off, model training, and generalization. Yanke Li
2. An introduction to data quality assessment and model evaluation metrics with a focus on explainability, robustness and generalization. Diego Paez
3. Tutorial for model training and feature selection methods in time-series. Mehdi Ejtehadi.
   1. Feature Selection Methods
   2. Building Classification Models

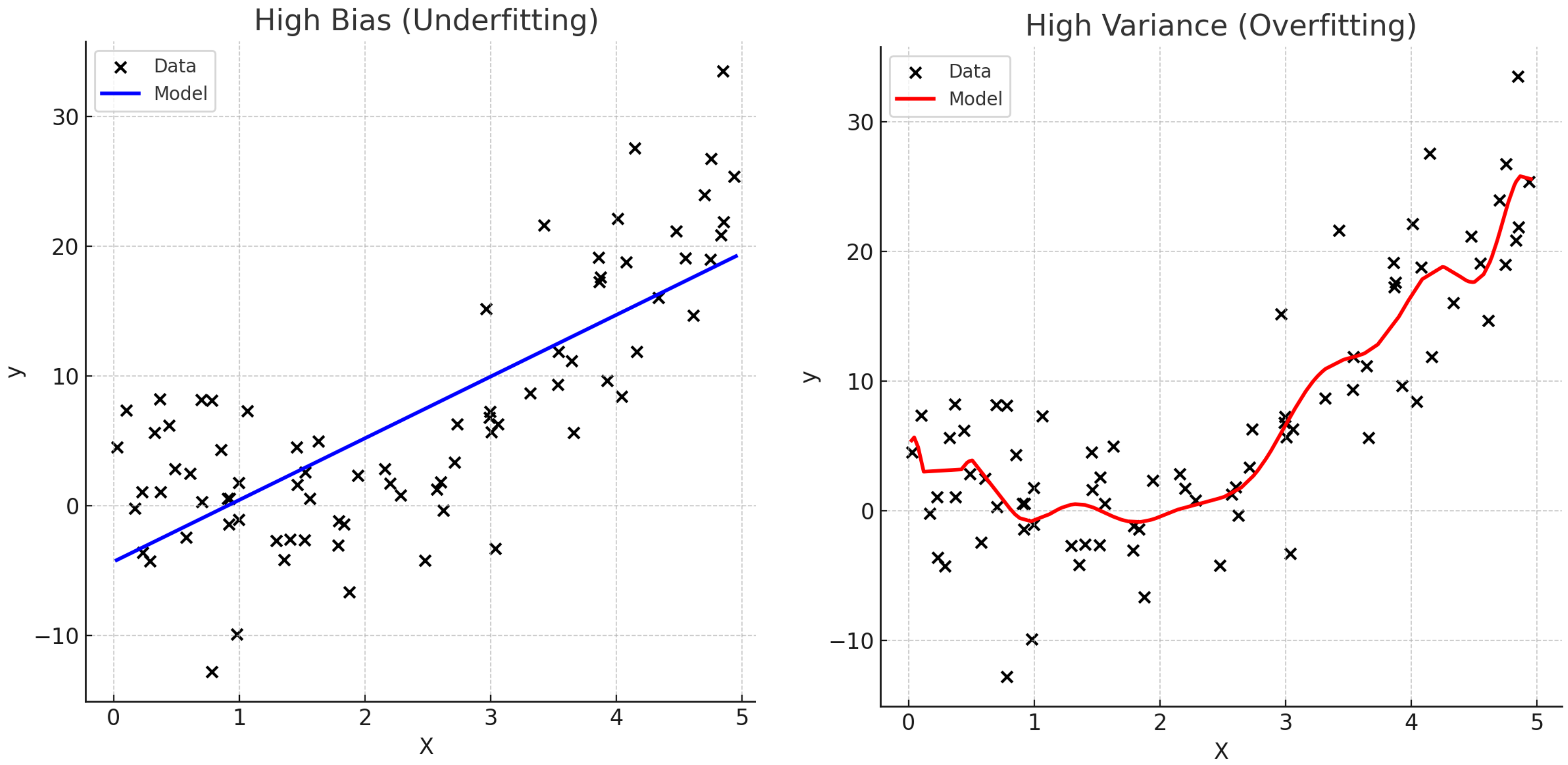
Model Quality: Fitness and Fairness

Dr Diego Paez
Mehdi Ejtehadi, Yanke Li, Bertram Fuchs



High Bias (Underfitting):

The left plot shows a linear model (blue line) that is too simple to capture the complexity of the data. The black dots (data points) are not well-fitted, resulting in high errors on both training and testing data.

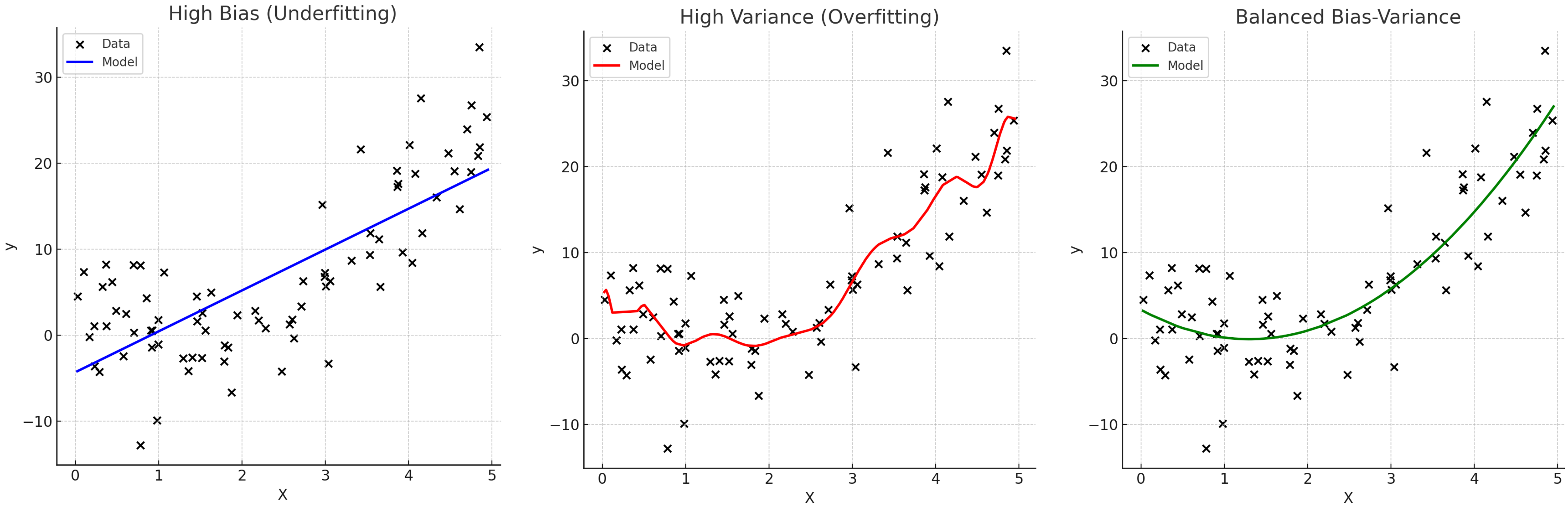


High Bias (Underfitting):

The left plot shows a linear model (blue line) that is too simple to capture the complexity of the data. The black dots (data points) are not well-fitted, resulting in high errors on both training and testing data.

High Variance (Overfitting):

The middle plot shows a highly complex polynomial model (red line) that fits the training data very closely, capturing the noise in the data. This results in low training error but a high error on new, unseen data.



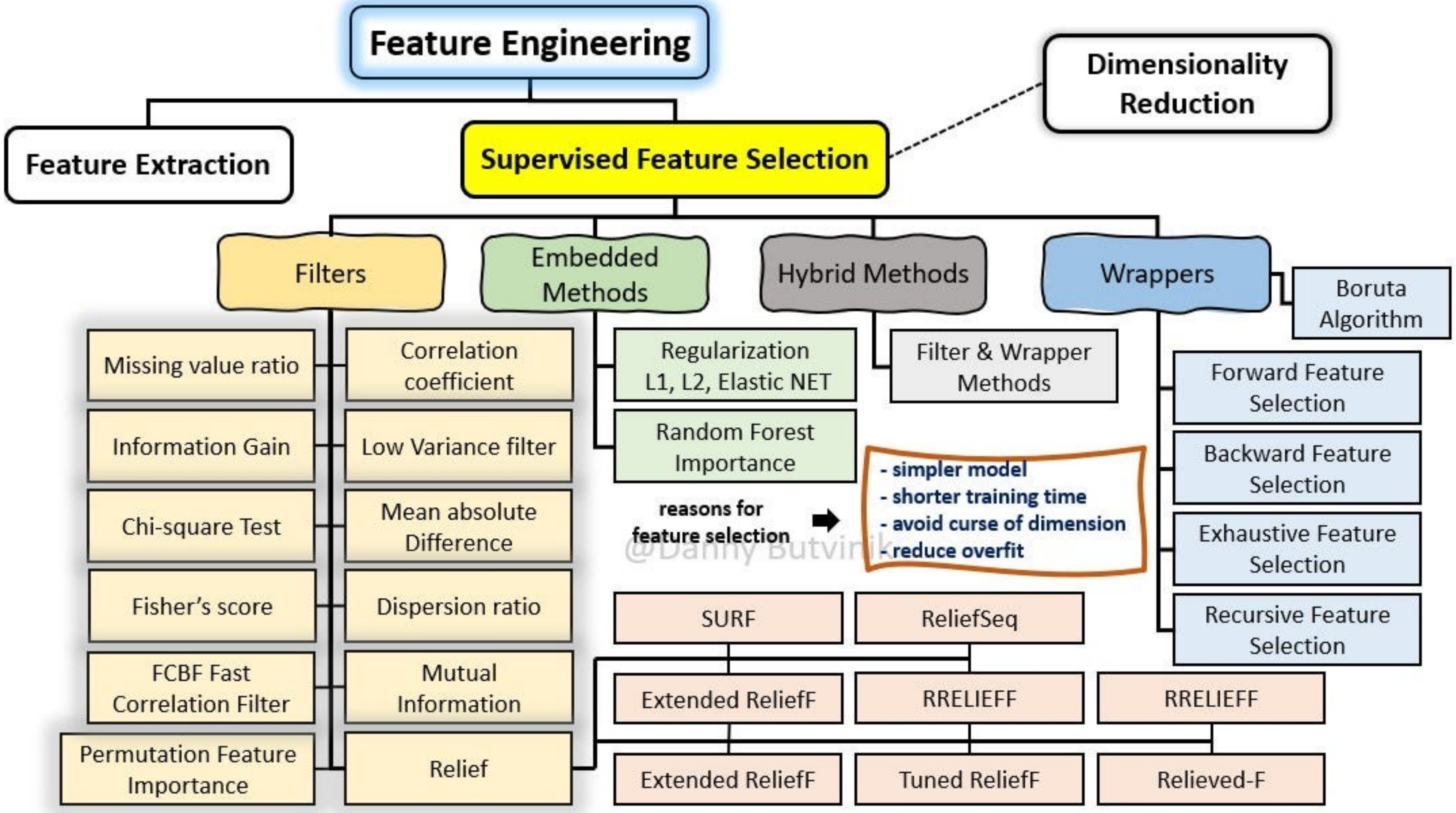
High Bias (Underfitting):

The left plot shows a linear model (blue line) that is too simple to capture the complexity of the data. The black dots (data points) are not well-fitted, resulting in high errors on both training and testing data.

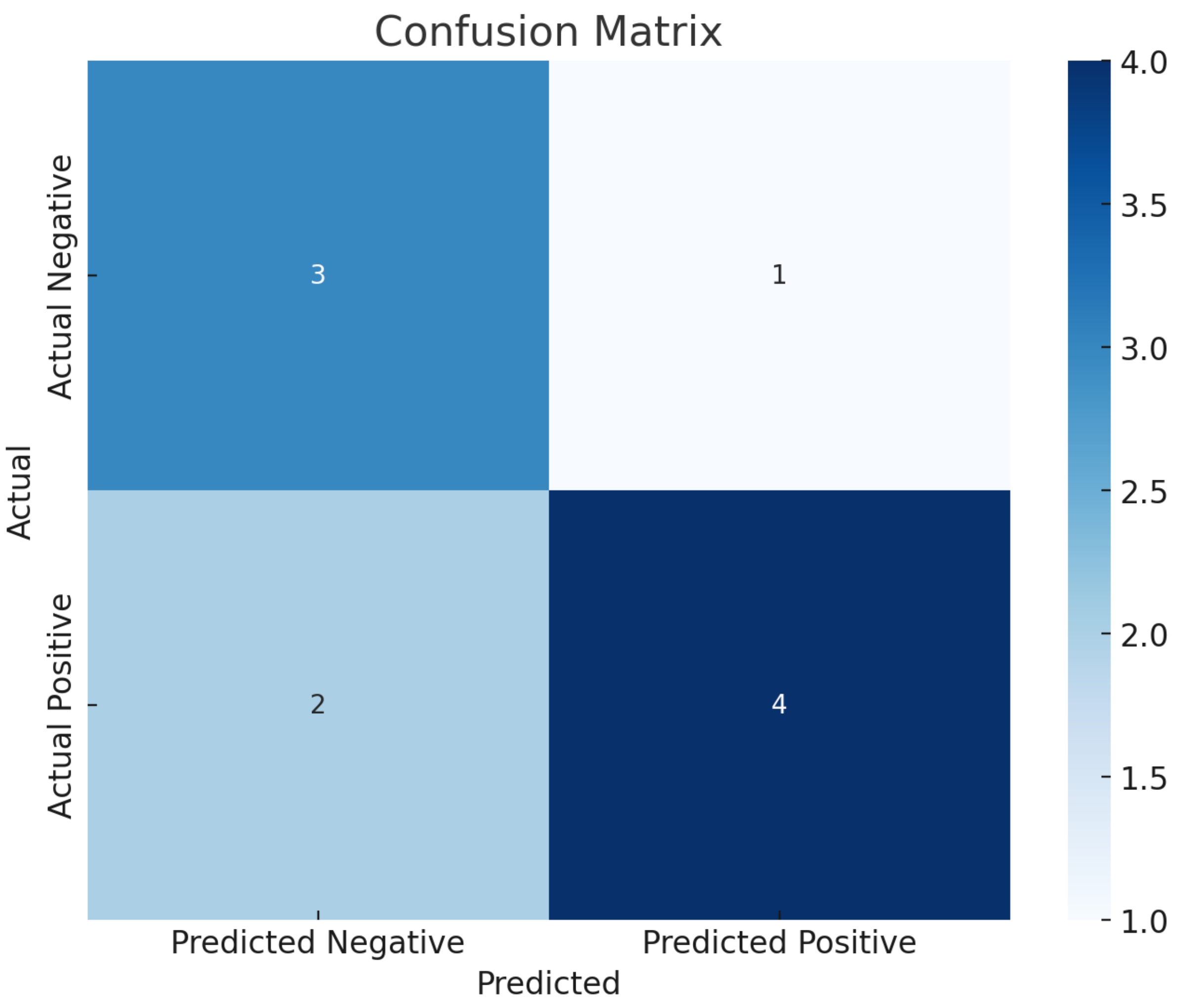
High Variance (Overfitting):

The middle plot shows a highly complex polynomial model (red line) that fits the training data very closely, capturing the noise in the data. This results in low training error but a high error on new, unseen data.

Feature Selection Taxonomy



Confusion Matrix



Accuracy

Percentage of correctly classified samples over all classes.

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Bias: unbalanced datasets won't be correctly evaluated

Balanced Accuracy:

$$Balanced\ Accuracy = \frac{Recall + Specificity}{2}$$

Recall or Sensitivity or TPR

A model's sensitivity can be found by answering the question,

"Of all the times when the true outcome class was 'Positive', how often did the model accurately predict that outcome?"

$$\frac{TP}{TP + FN}$$

Bias: Skewing the score towards positive examples.
E.g. - Everything as positive (no FN) will give a full score.

Precision

A model's precision can be found by answering the question, "Of all the times when the model predicted that a record would belong to the 'Positive' outcome class, how often was it correct?"

$$\frac{TP}{TP + FP}$$

Bias: Negative labels are not accounted.

F1-Score

The model's F1 score combines precision and recall into a single metric by taking their harmonic mean, and is particularly useful when comparing models.

When a model is adjusted so that false positives become rarer, that makes it less likely to identify any record as belonging to the positive outcome class. This brings precision higher, but it generally brings recall lower.

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Micro-Averaged F1 Score
Airplane	2	1	1	$\frac{\mathrm{TP}}{\mathrm{TP}+\frac{1}{2}(\mathrm{FP}+\mathrm{FN})} = \frac{6}{6+\frac{1}{2}(4+4)}$ = **0.60**
Boat	1	3	0	
Car	3	0	3	
TOTAL	**6**	**4**	**4**	

$$\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * TP}{2 * TP + FP + FN}$$

Specificity or TNR

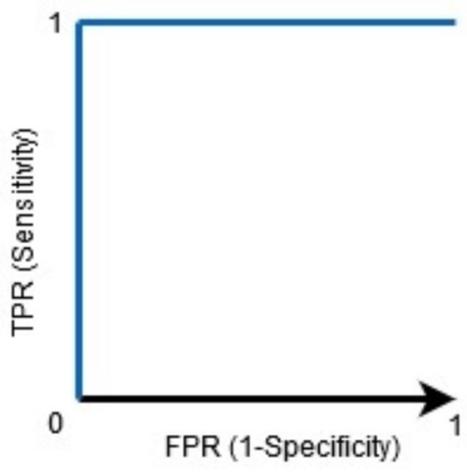
A model's specificity can be found by answering the question,

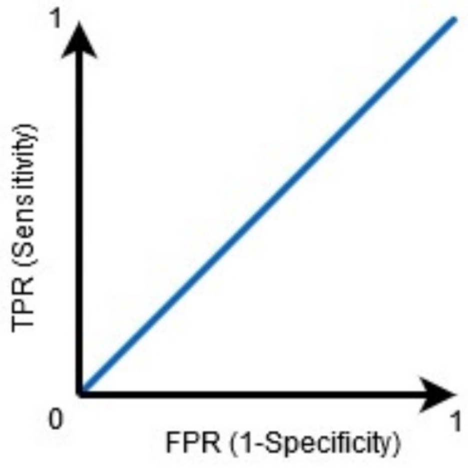
"Of all the times when the true outcome class was 'Negative', how often did the model accurately predict that outcome?"

$$\frac{TN}{TN + FP}$$

Precision-Recall Curve

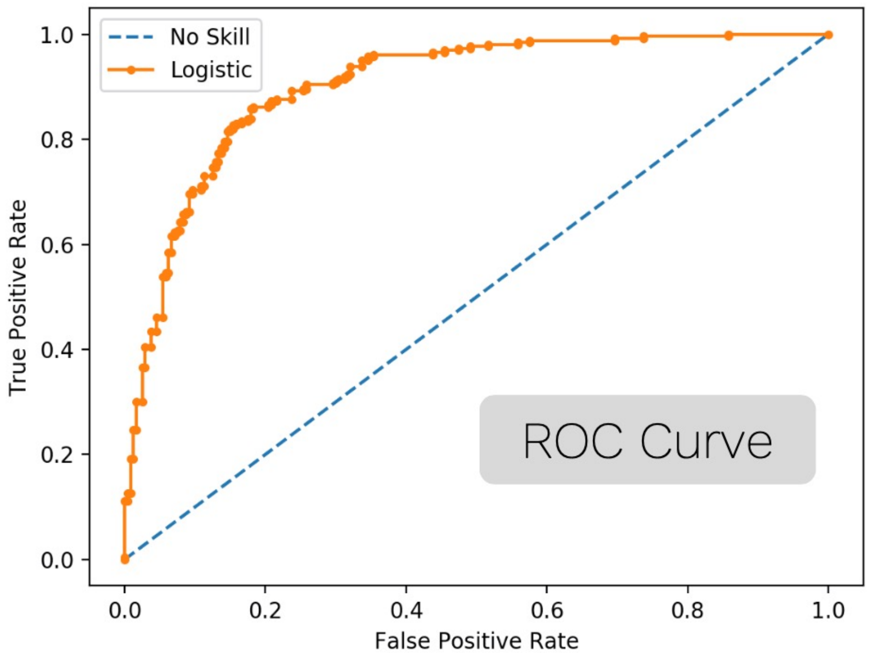
Definition: Precision is the proportion of true positive predictions out of all positive predictions made by the model. It answers the question, "Of all instances the model predicted as positive, how many were actually positive?"





Precision-Recall Curve

Definition: Precision is the proportion of true positive predictions out of all positive predictions made by the model. It answers the question, "Of all instances the model predicted as positive, how many were actually positive?"



Source: https://medium.com/swlh/how-to-remember-all-these-classification-concepts-forever-761c065be33

Summary

Predicted Class

Actual Class		**Positive**	**Negative**	
	Positive	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	**Negative Predictive Value** $\frac{TN}{(TN + FN)}$	**Accuracy** $\frac{TP + TN}{(TP + TN + FP + FN)}$

Source: https://medium.com/swlh/how-to-remember-all-these-classification-concepts-forever-761c065be33

Summary

•Accuracy reflects the overall accuracy of the model's predictions.

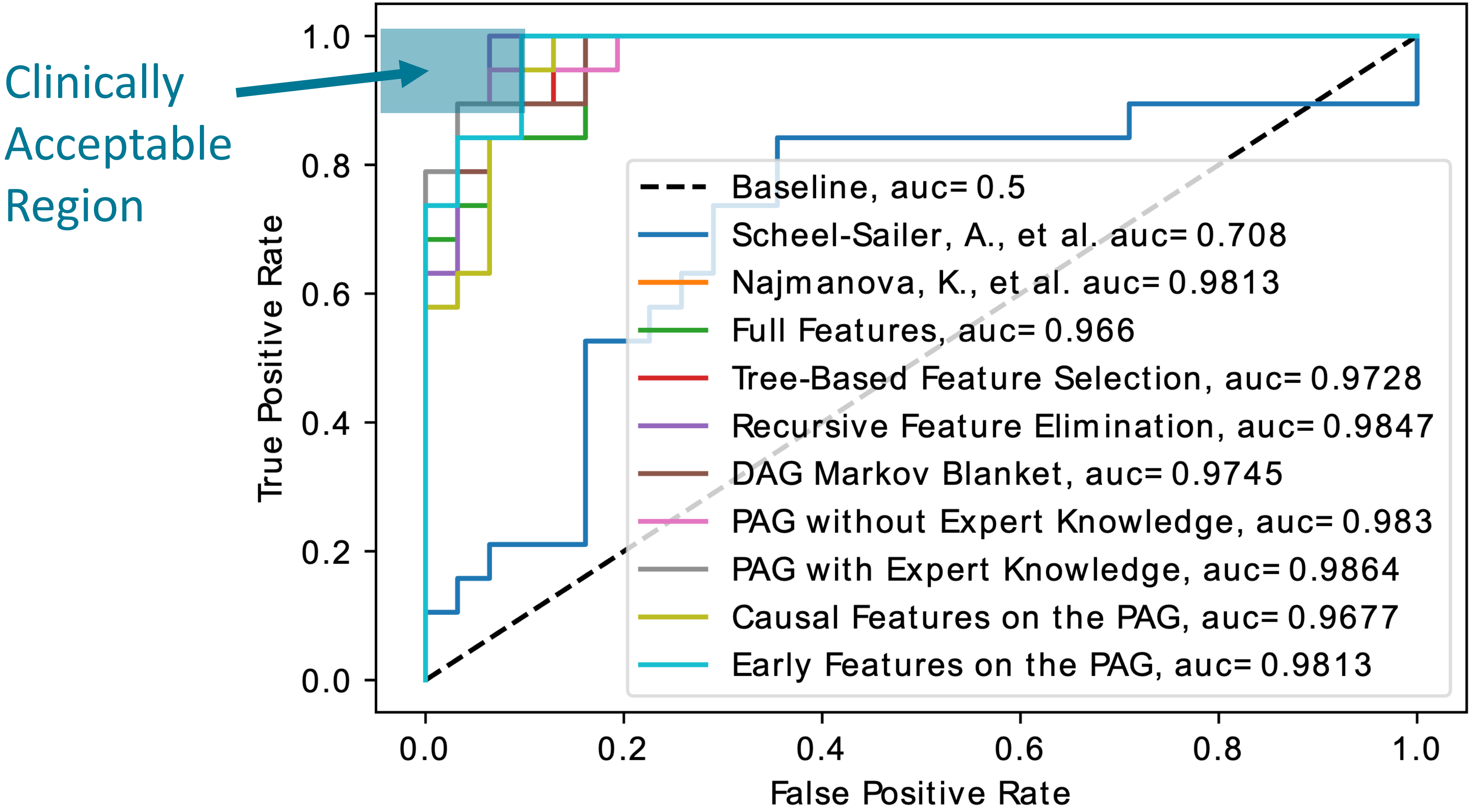
•Precision measures the proportion of positive predictions that were actually correct.

•Recall, also known as sensitivity or true positive rate, captures the proportion of actual positive samples that were correctly predicted by the model.

•False positive rate represents the proportion of actual negative samples that were incorrectly predicted as positive by the model.

•False negative rate reflects the proportion of actual positive samples that were incorrectly predicted as negative by the model.

Performance of Predicting Pressure Injuries in Hospital Stays for SCI Using Graphical Models



Y. Li, A. Scheel-Sailer, J. Pannek, R. Riener, and D. Paez-Granados. "Mixed-Variable Graphical Modelling for Medical Record-Based Prediction Models: Learning Pressure Injury Onset from Spinal Cord Injuries". In: Under review - (2023).

Tutorial: