Tutorial 5 - Performance Metrics

Kevin Dick, PhDc Biomedical Engineering

Carleton University

Friday 9th October, 2020

Disclaimer: Recorded Tutorials will be Publicly Posted

Goal: to create a companion series of applied machine learning tutorials for the 100MLB text, these tutorials will be publicly posted as a YouTube playlist.

Privacy Preservation:

- Ask questions in the chat[1]
- Keep video off

Note: If the above *hinders your ability to learn* $\wedge$ *violates your privacy*, please let me/Dr. Green know ASAP and video will be post-processed accordingly.

[1] I encourage unmuted/voice-based questions at any time, but know that this isn't explicitly privacy-preserving

ML Weekly

Recent news events from the ML community

ML Weekly

1. **(ML)** NeurIPS 2020 Accepted Papers

ML Weekly

1. **(ML)** NeurIPS 2020 Accepted Papers
2. **(Vision)** "Farewell Convolutions": Anonymous ICLR 2021 Paper Uses Transformers for Image Recognition at Scale

AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

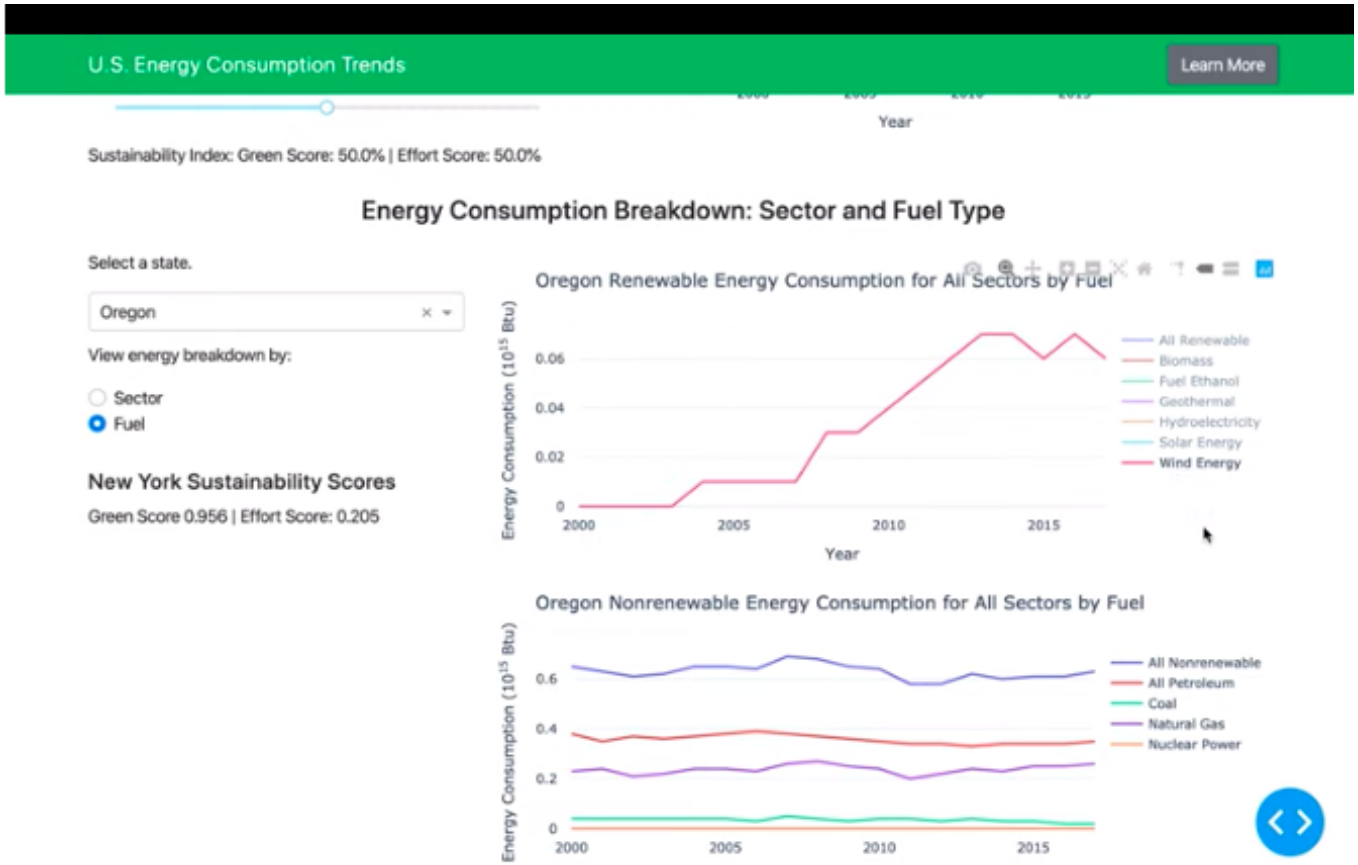
Anonymous authors
Paper under double-blind review

ABSTRACT

While the Transformer architecture has become the de-facto standard for natural language processing tasks, its applications to computer vision remain limited. In vision, attention is either applied in conjunction with convolutional networks, or used to replace certain components of convolutional networks while keeping their overall structure in place. We show that this reliance on CNNs is not necessary and a pure transformer can perform very well on image classification tasks when applied directly to sequences of image patches. When pre-trained on large amounts of data and transferred to multiple recognition benchmarks (ImageNet, CIFAR-100, VTAB, etc.), Vision Transformer attain excellent results compared to state-of-the-art convolutional networks while requiring substantially fewer computational resources to train.

ML Weekly

1. **(ML)** NeurIPS 2020 Accepted Papers
2. **(Vision)** "Farewell Convolutions": Anonymous ICLR 2021 Paper Uses Transformers for Image Recognition at Scale
3. **(DS)** Plotly Visualization



ML Weekly

1. **(ML)** NeurIPS 2020 Accepted Papers
2. **(Vision)** "Farewell Convolutions": Anonymous ICLR 2021 Paper Uses Transformers for Image Recognition at Scale
3. **(DS)** Plotly Visualization
4. **(NLP)** Reddit: thegentlemetre ("GPT-3 Bot Spends a Week Replying on Reddit, Starts Talking About the Illuminati")

Tutorial Intuition

Building an Intuition for the Concepts of this Tutorial

How Good is my Model?

Welcome to the *metric zoo*!

There exists several dozens of model evaluation metrics and it is critical that one pairs the **right metric** for the **right problem**.

How Good is my Model?

Welcome to the *metric zoo*!

There exists several dozens of model evaluation metrics and it is critical that one pairs the **right metric** for the **right problem**.

- **Classification-Type:** accuracy, precision, recall, specificity, F1-score, ROC, AUC, ...
- **Regression-Type:** MSE, MAE, RMSE, ...

How Good is my Model?

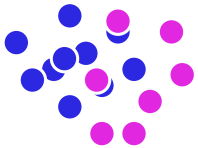
Welcome to the *metric zoo*!

There exists several dozens of model evaluation metrics and it is critical that one pairs the **right metric** for the **right problem**.

- **Classification-Type:** accuracy, precision, recall, specificity, F1-score, ROC, AUC, ...
- **Regression-Type:** MSE, MAE, RMSE, ...
- Rank-Aware Metrics
- Statistical-Type Metrics
- Computer Vision Metrics
- NLP Metrics
- Deep Learning-Related Metrics

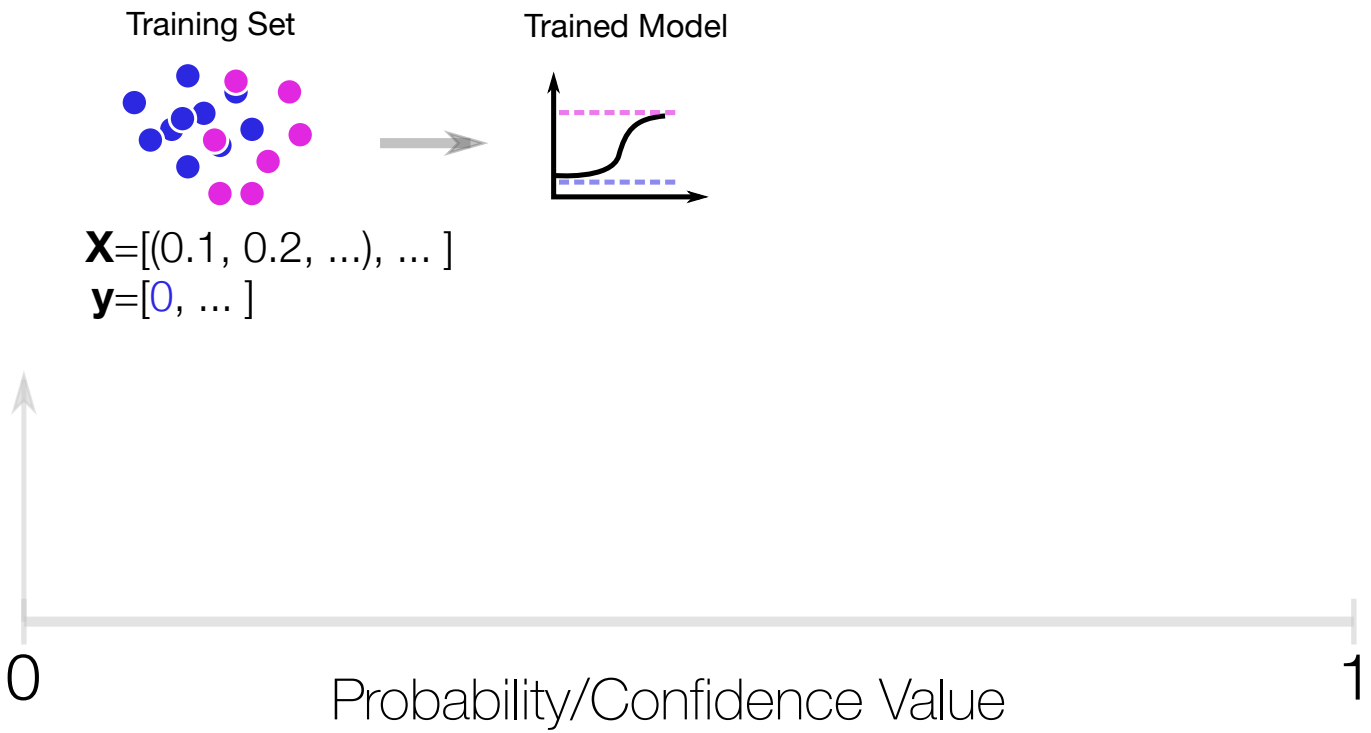
How Good is my Model?

Training Set

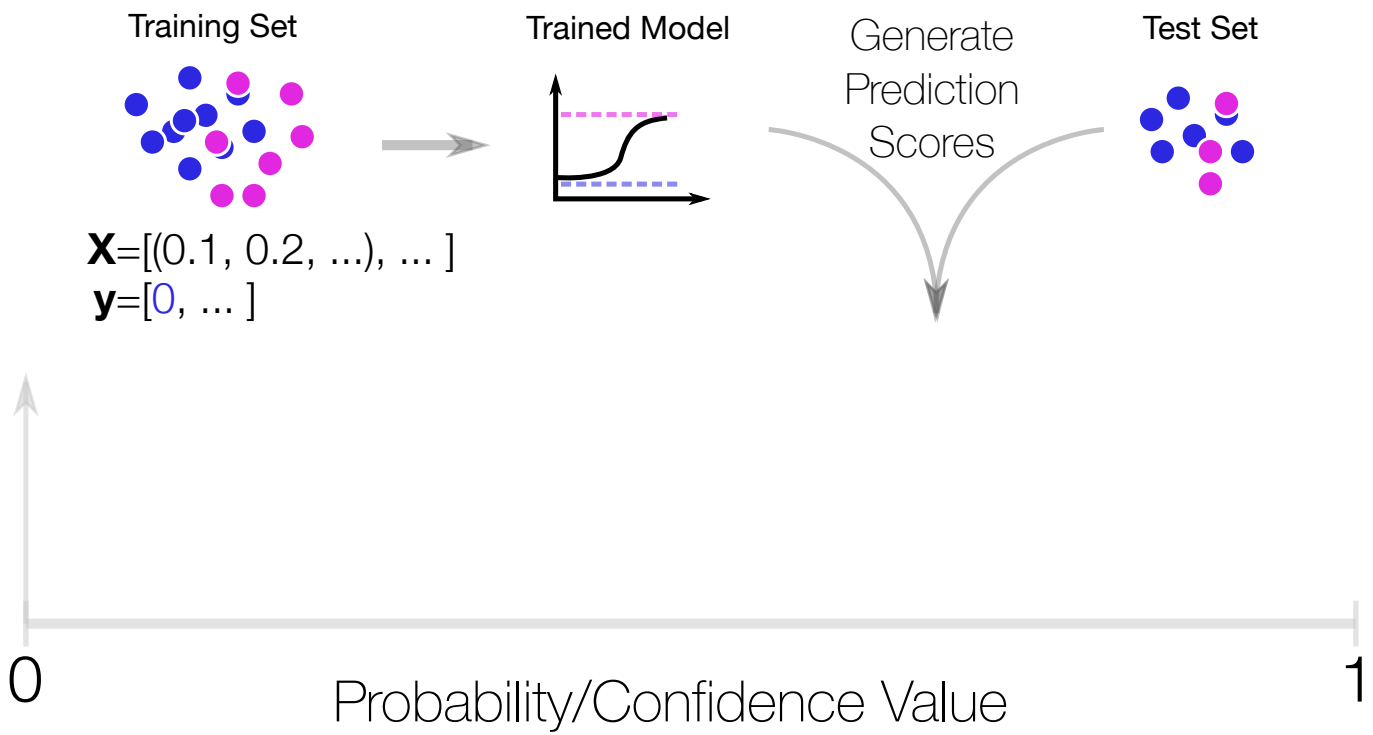


X=[(0.1, 0.2, ...), ...]
y=[0, ...]

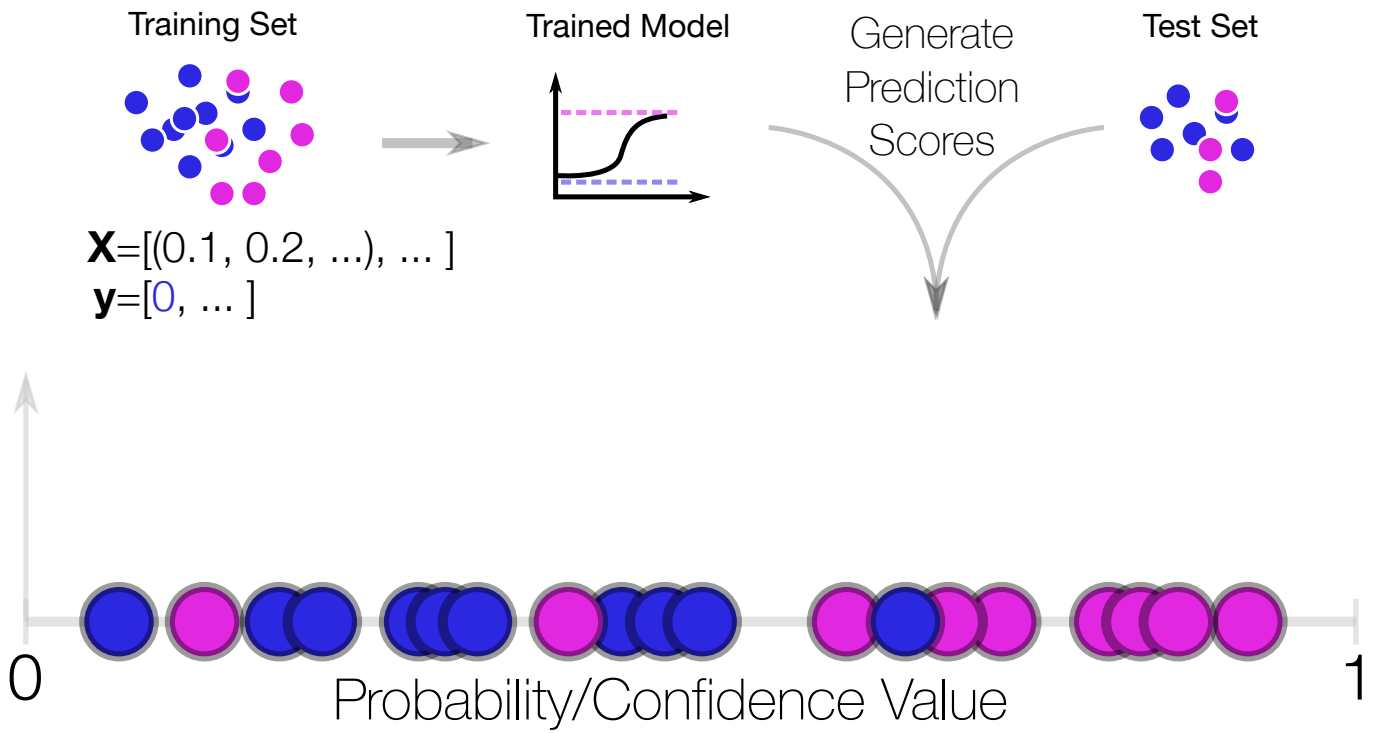
How Good is my Model?



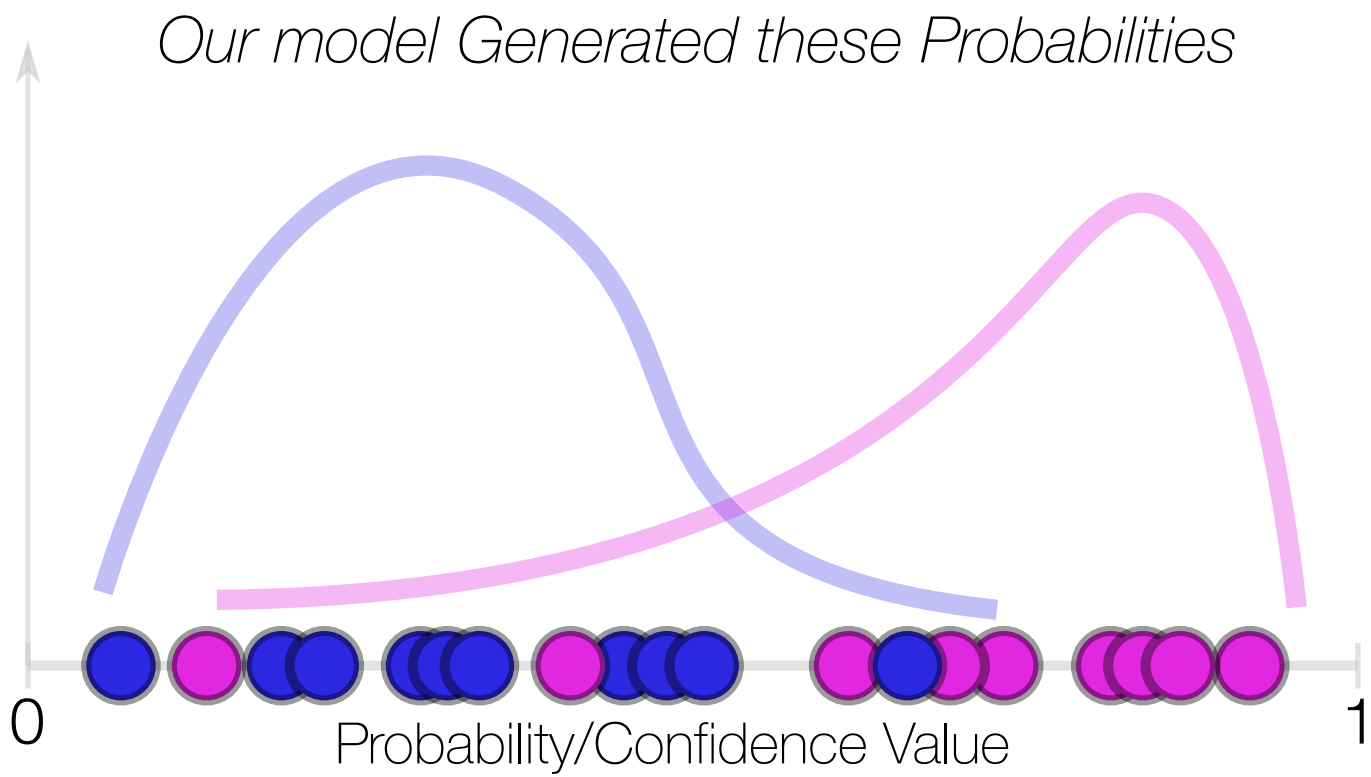
How Good is my Model?



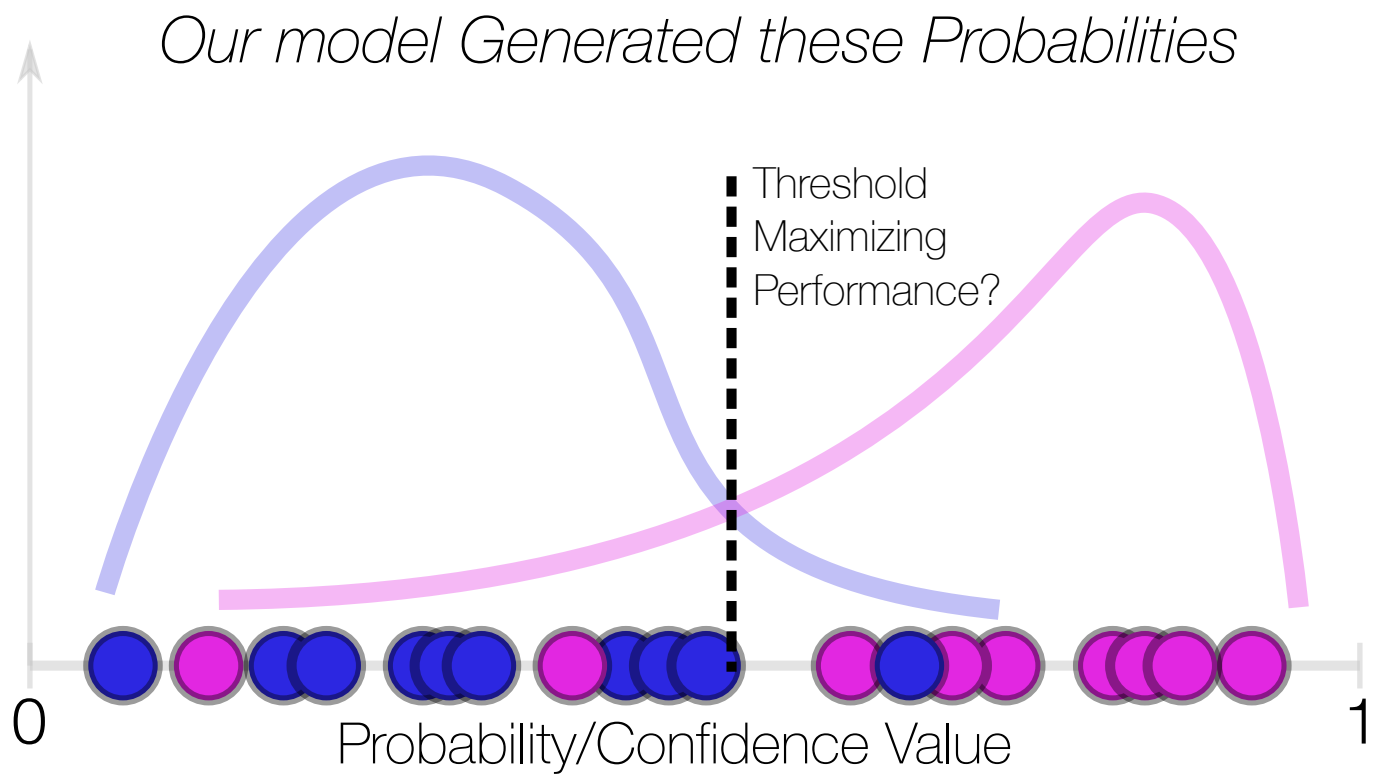
How Good is my Model?



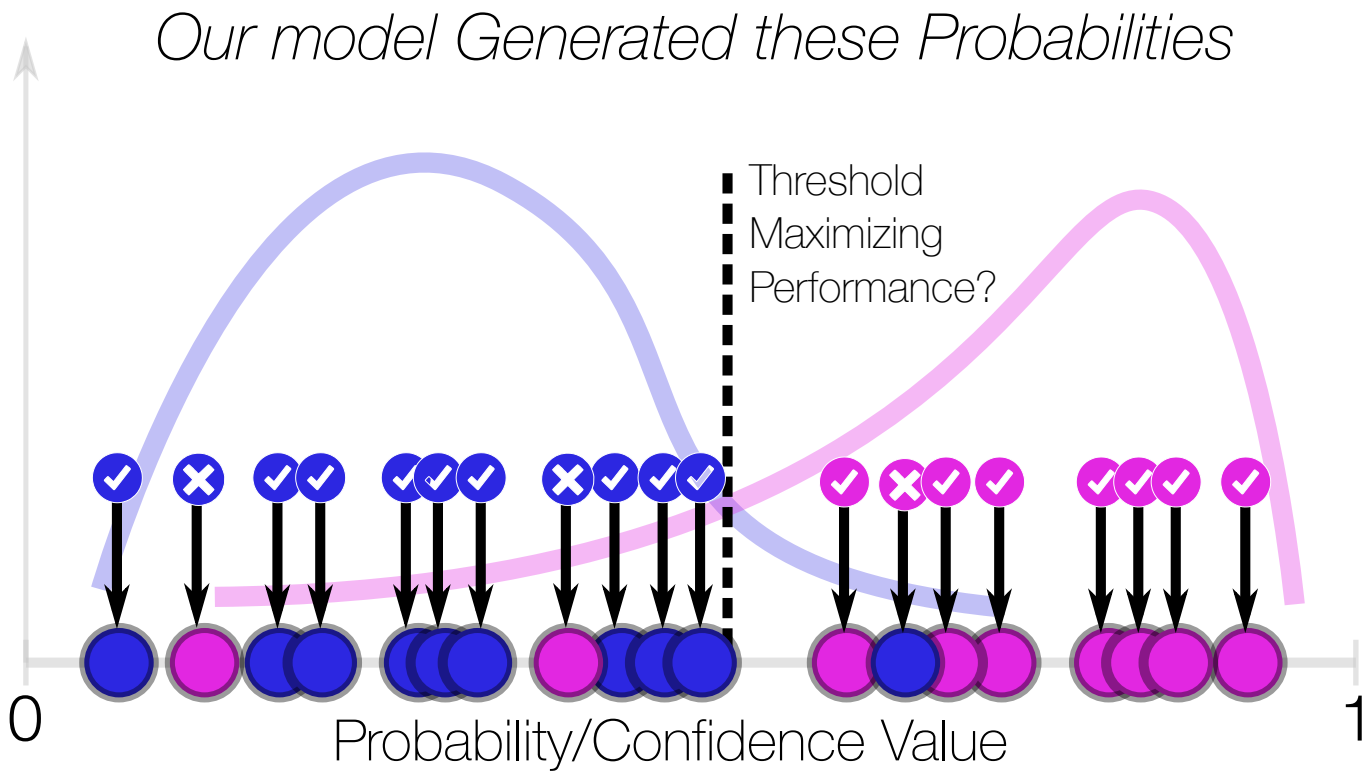
How Good is my Model?



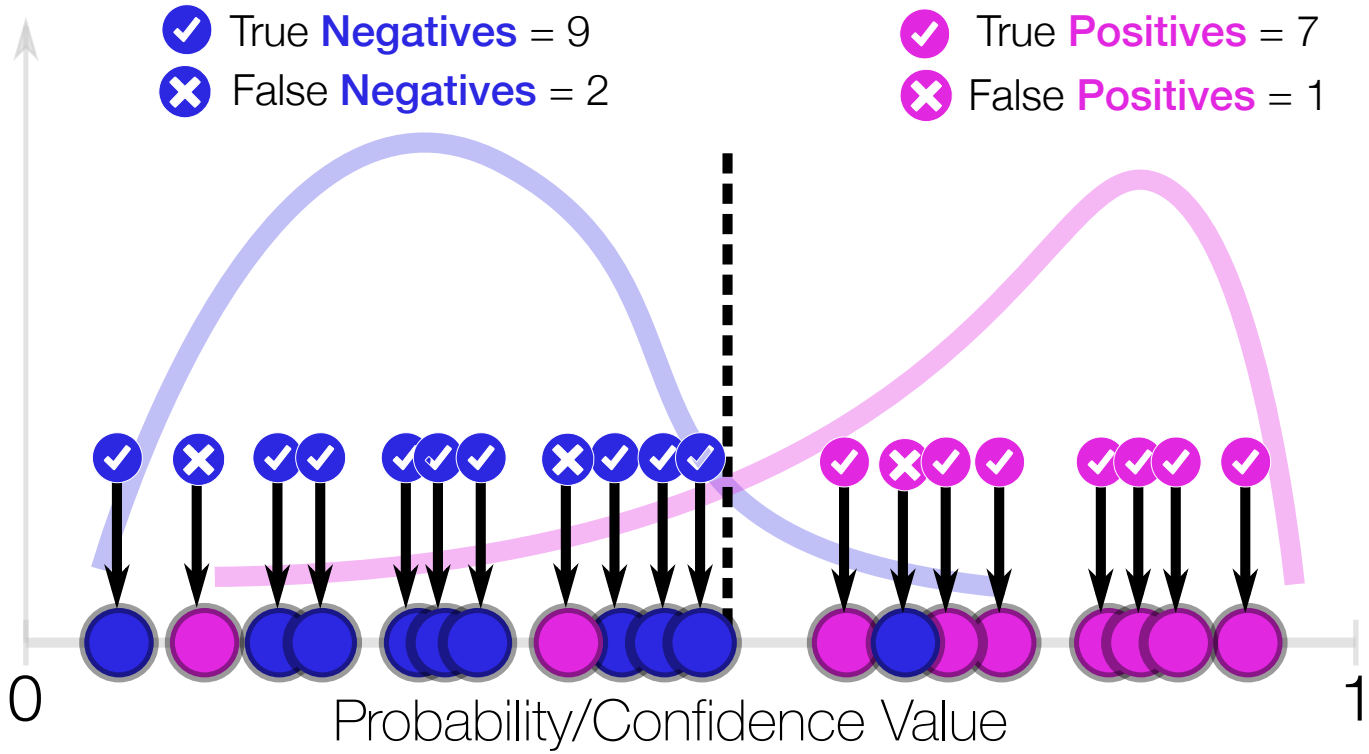
How Good is my Model?



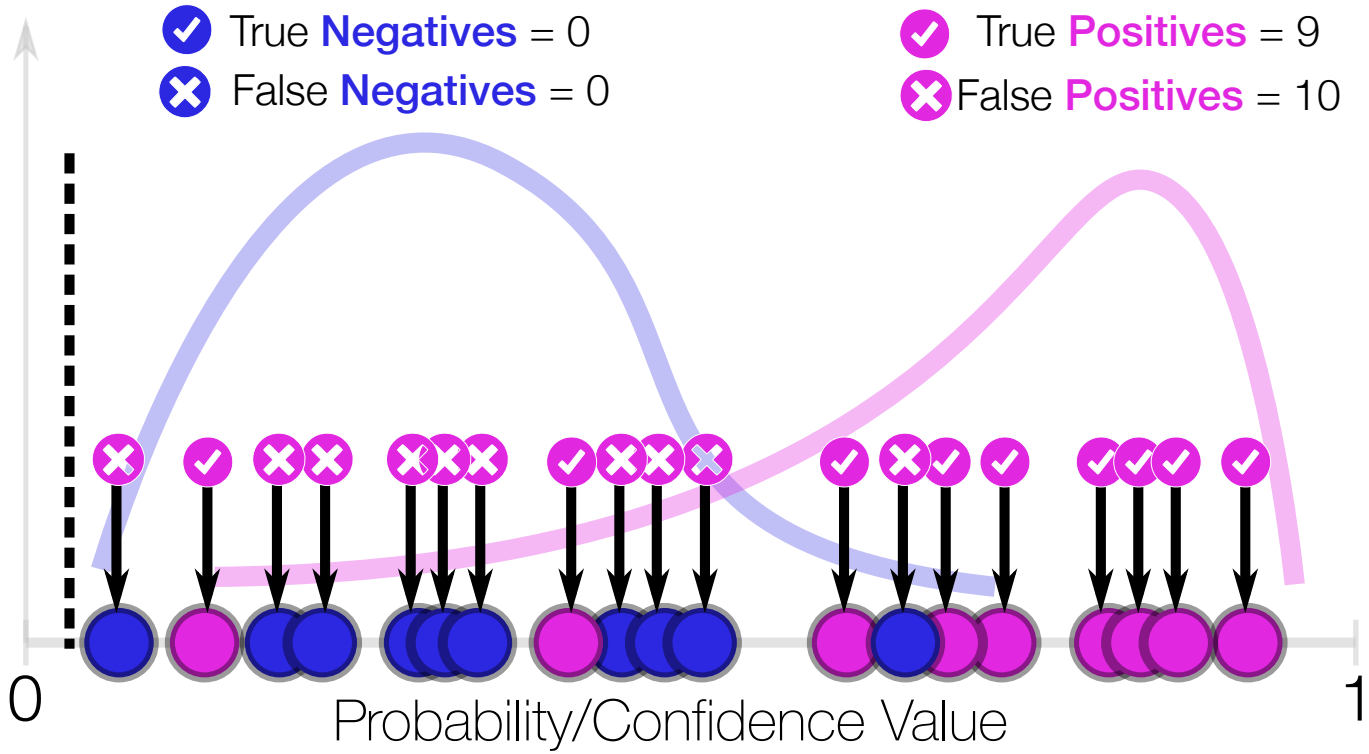
How Good is my Model?



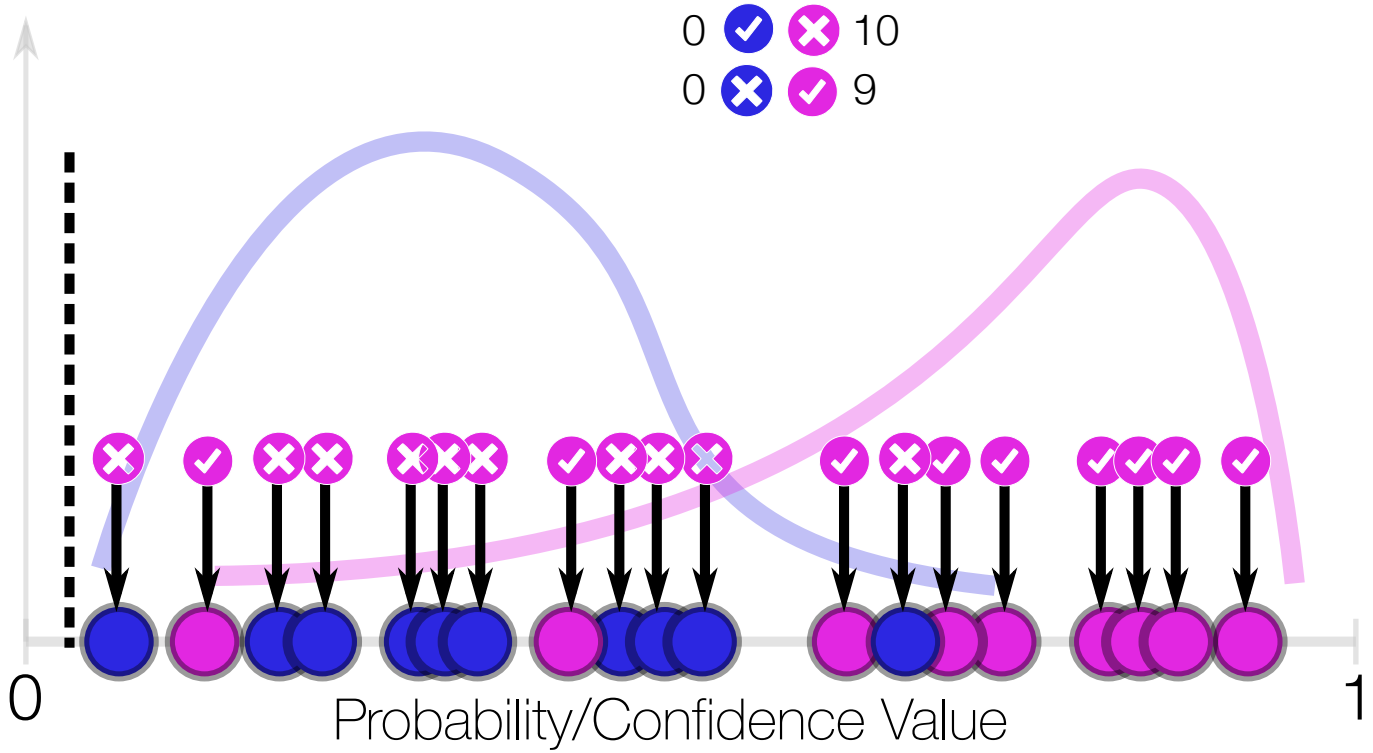
How Good is my Model?



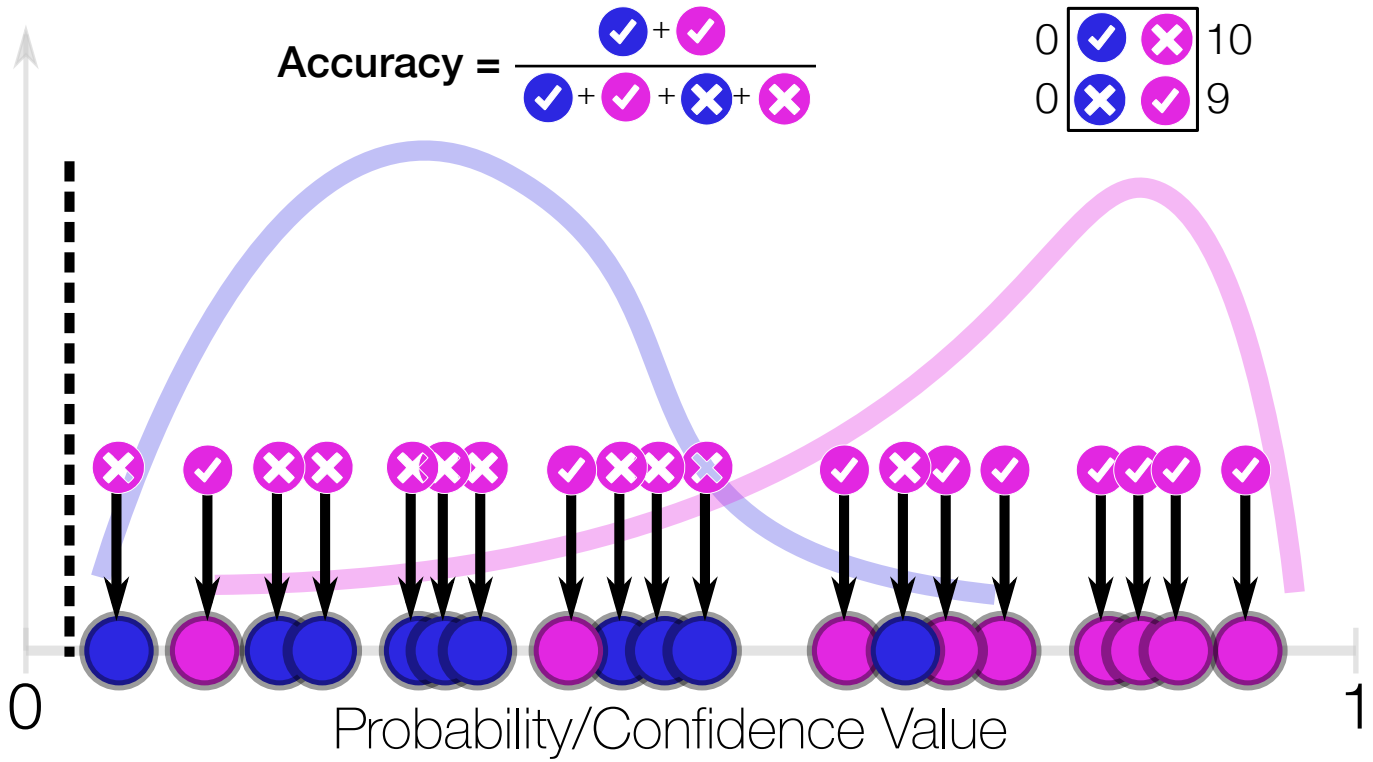
How Good is my Model?



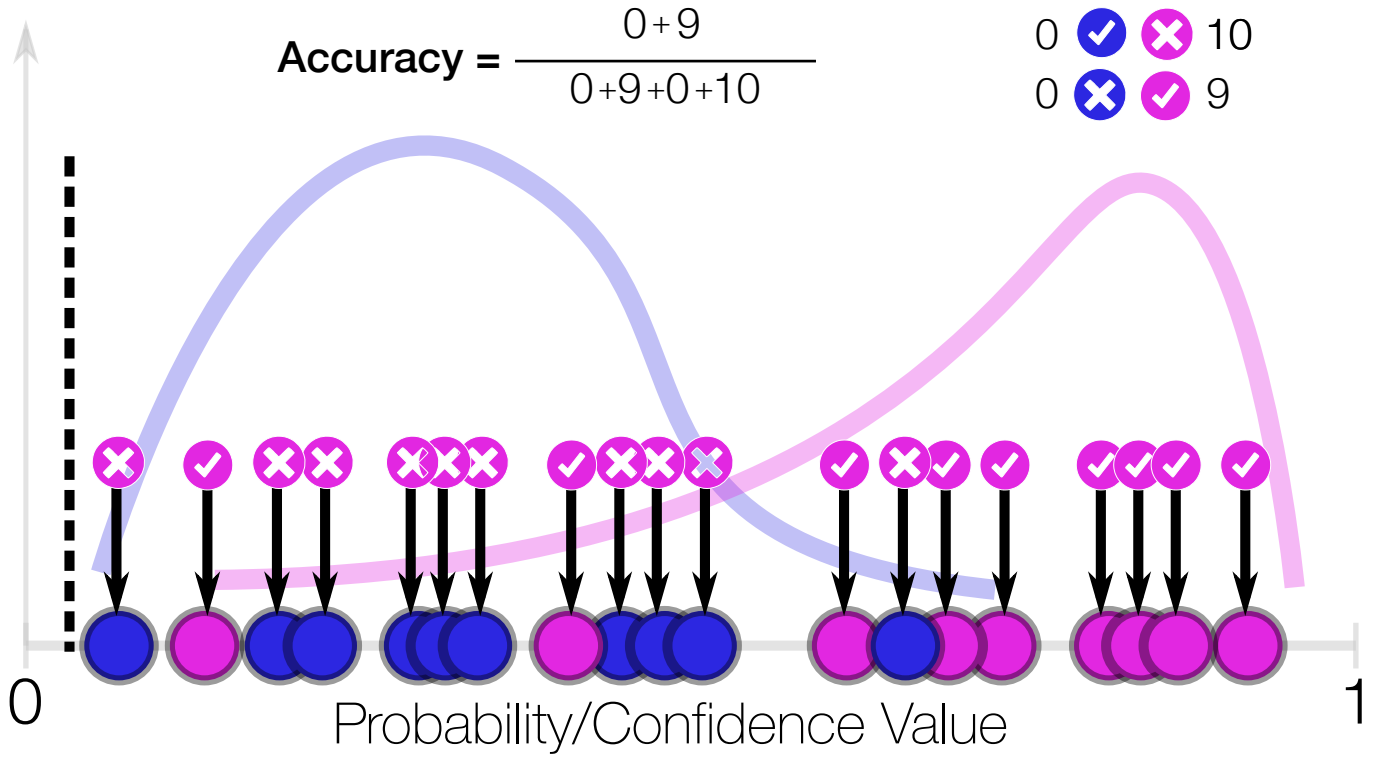
How Good is my Model?



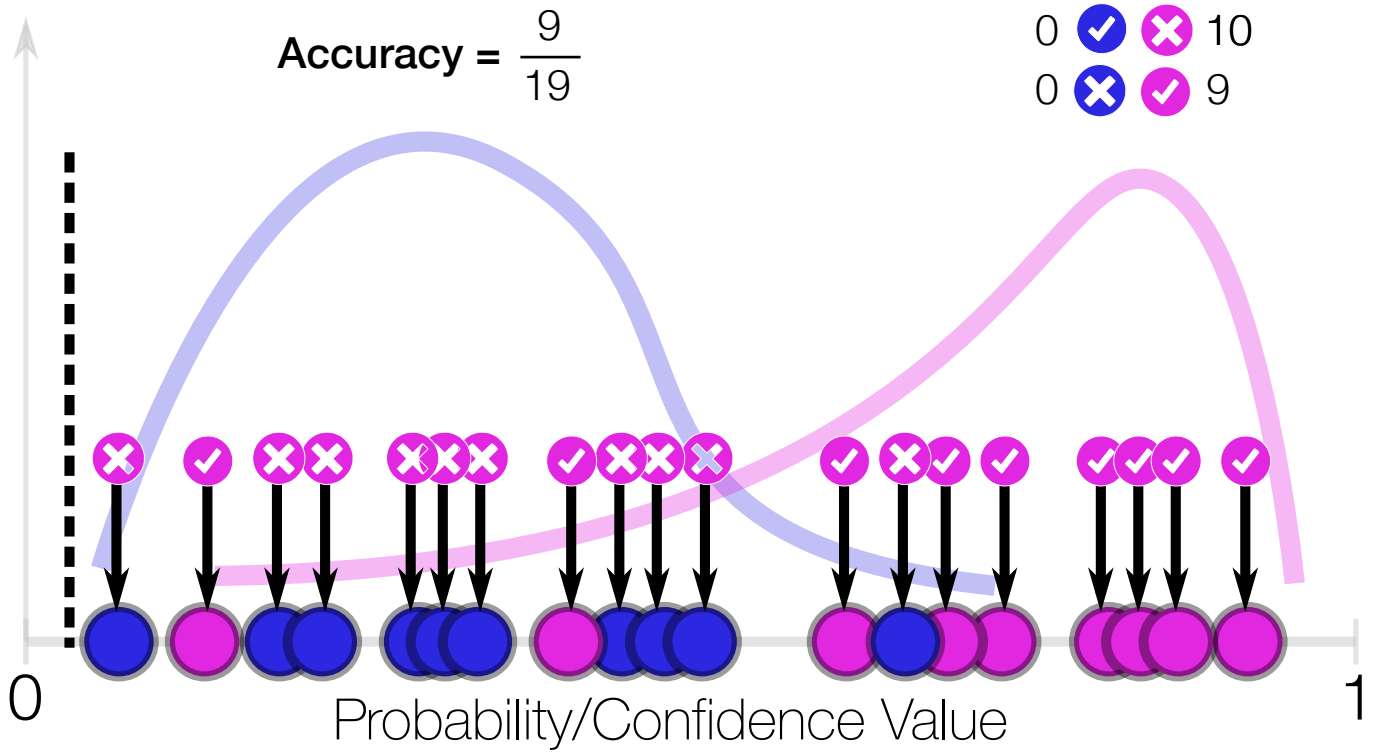
How Good is my Model?



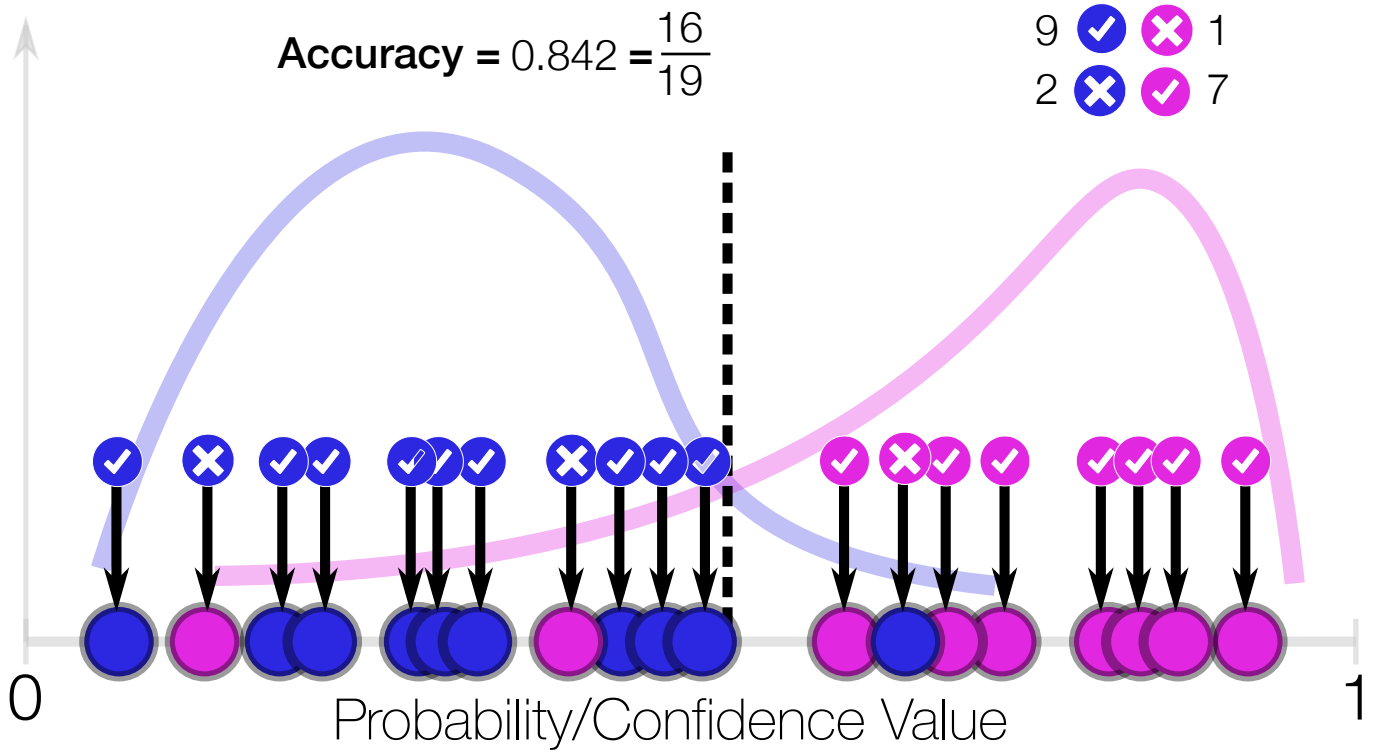
How Good is my Model?



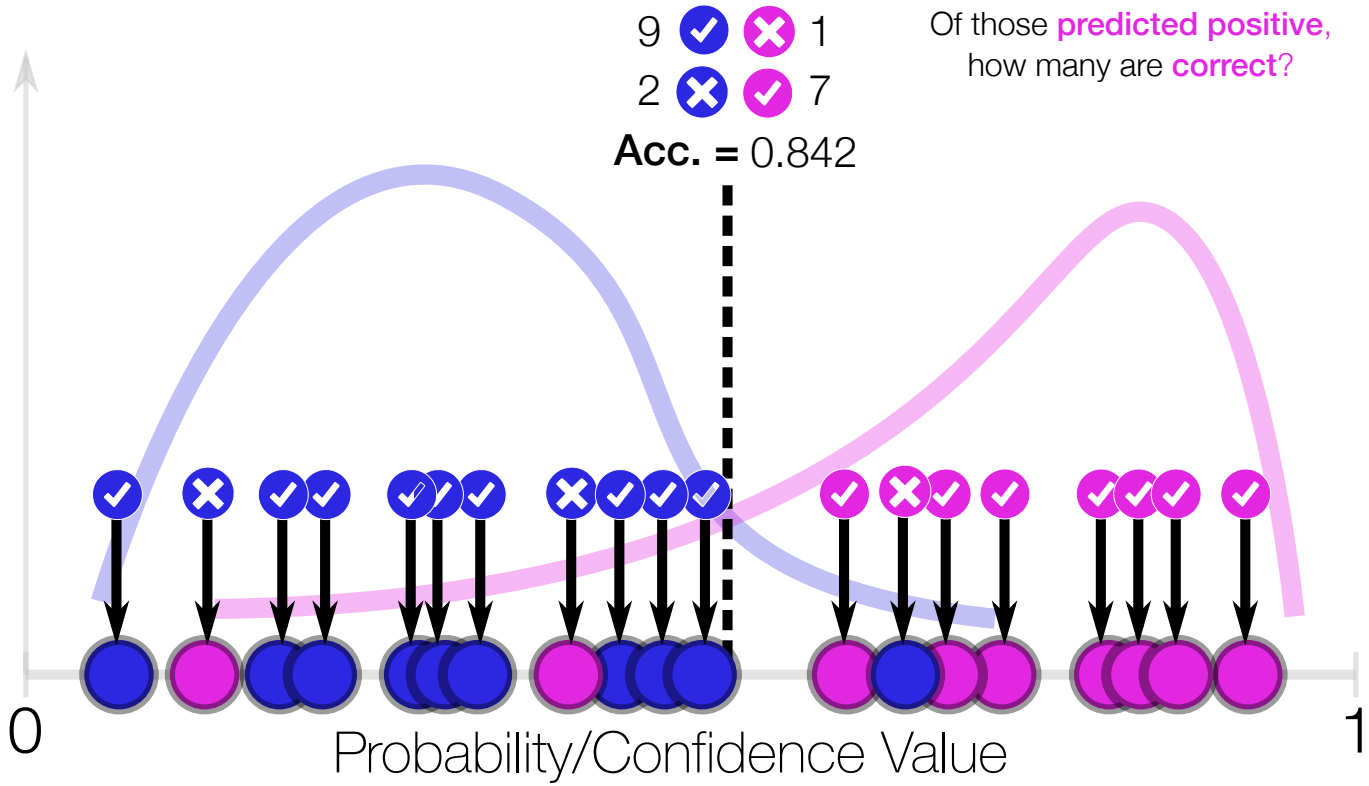
How Good is my Model?



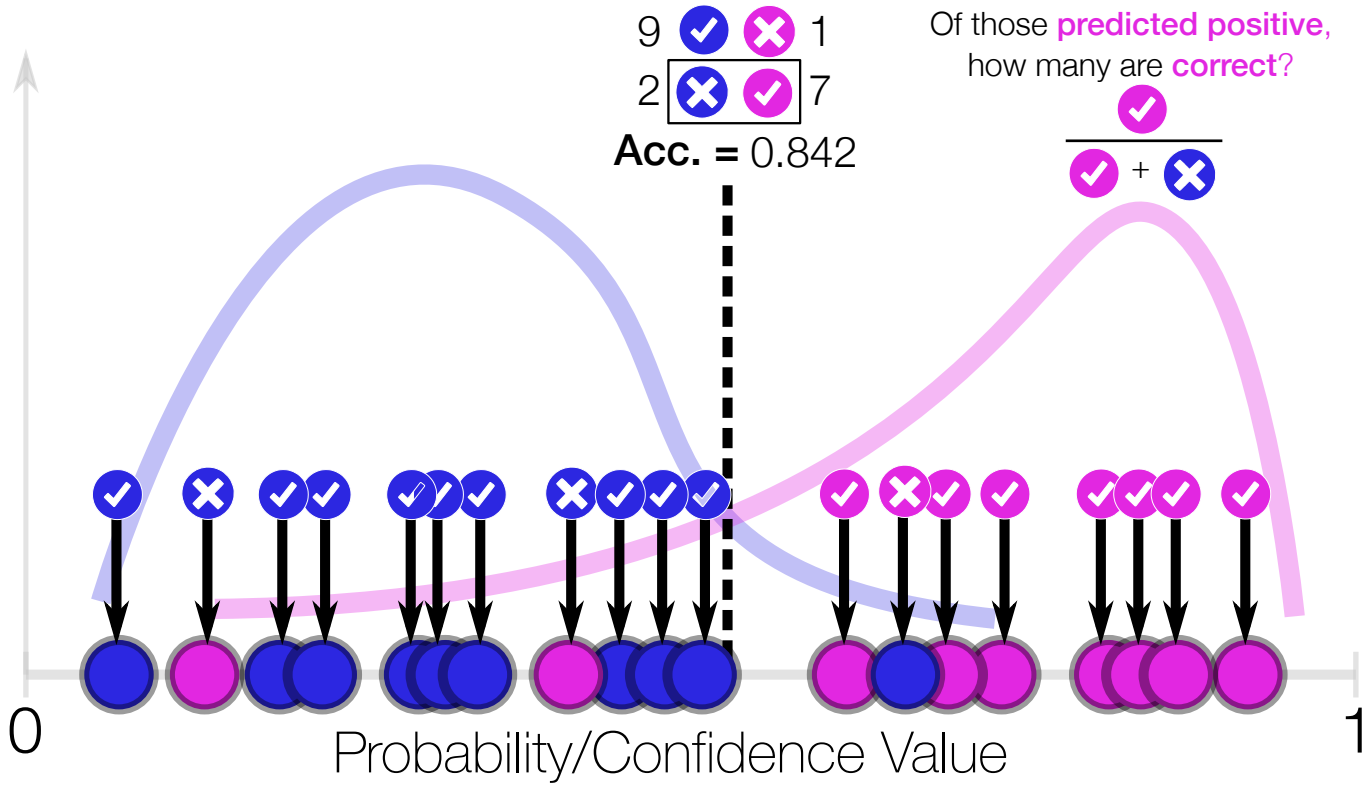
How Good is my Model?



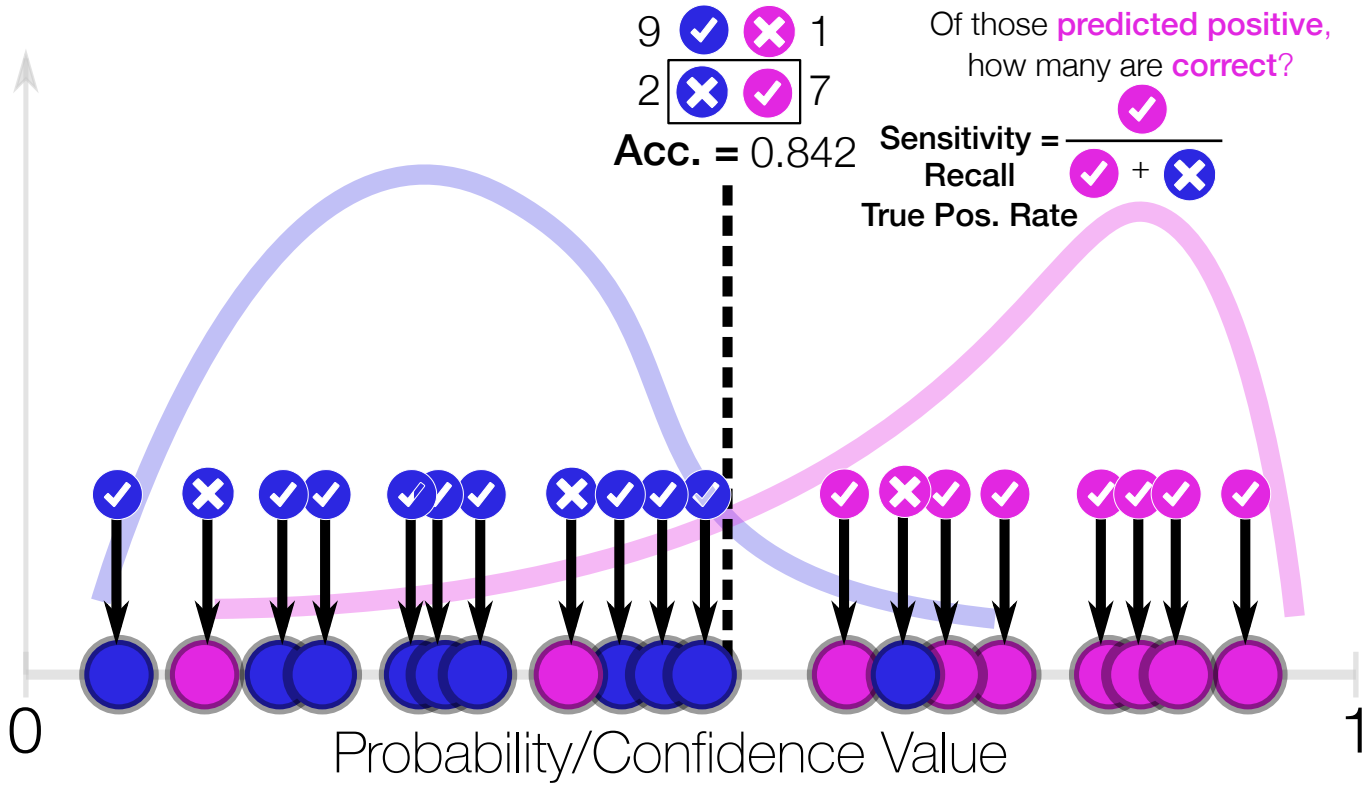
How Good is my Model?



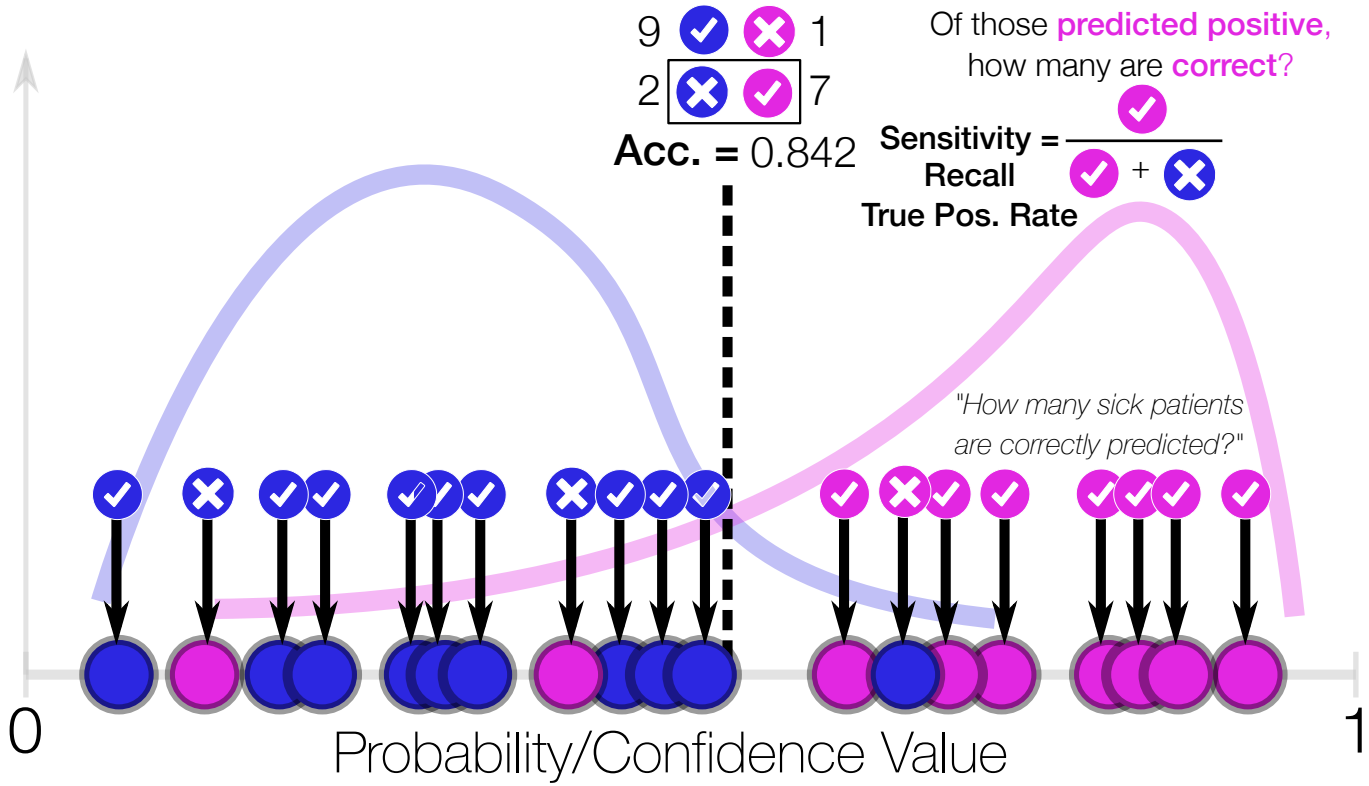
How Good is my Model?



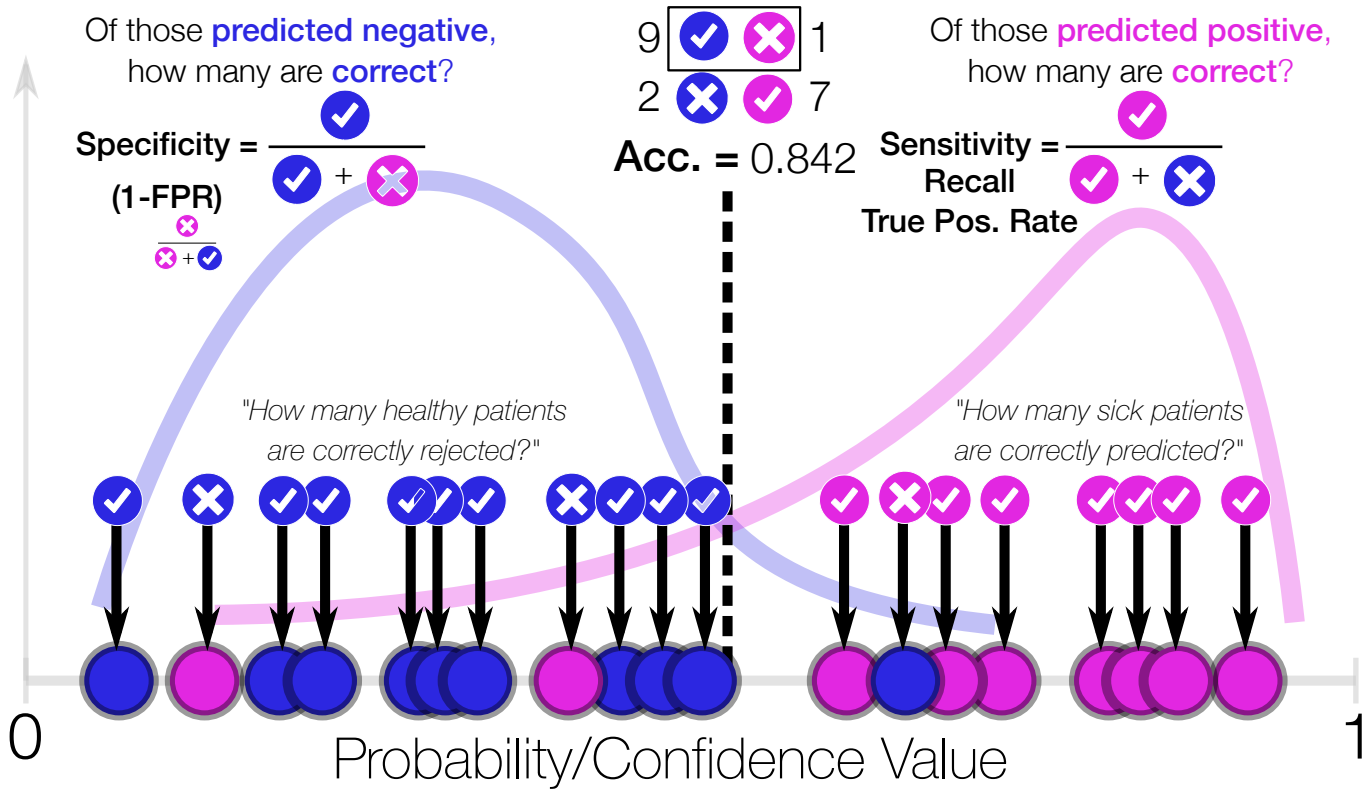
How Good is my Model?



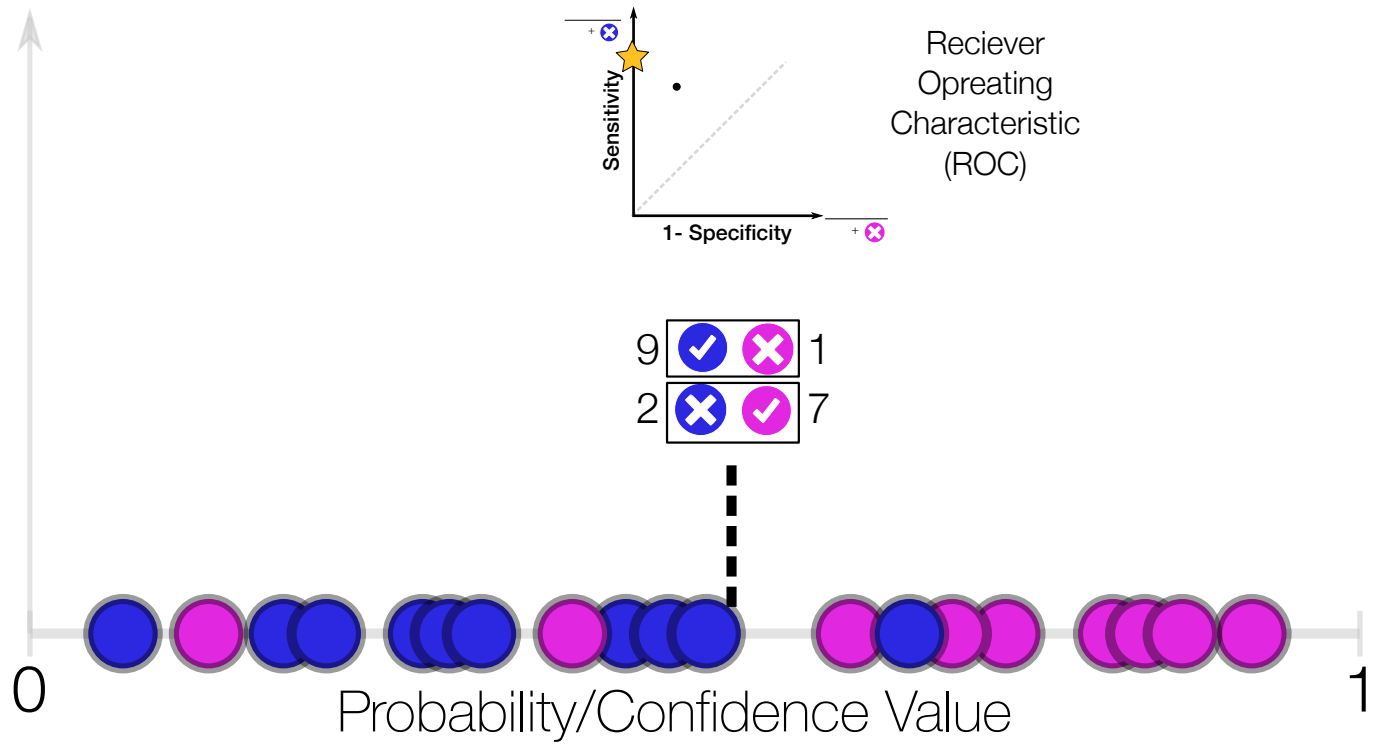
How Good is my Model?



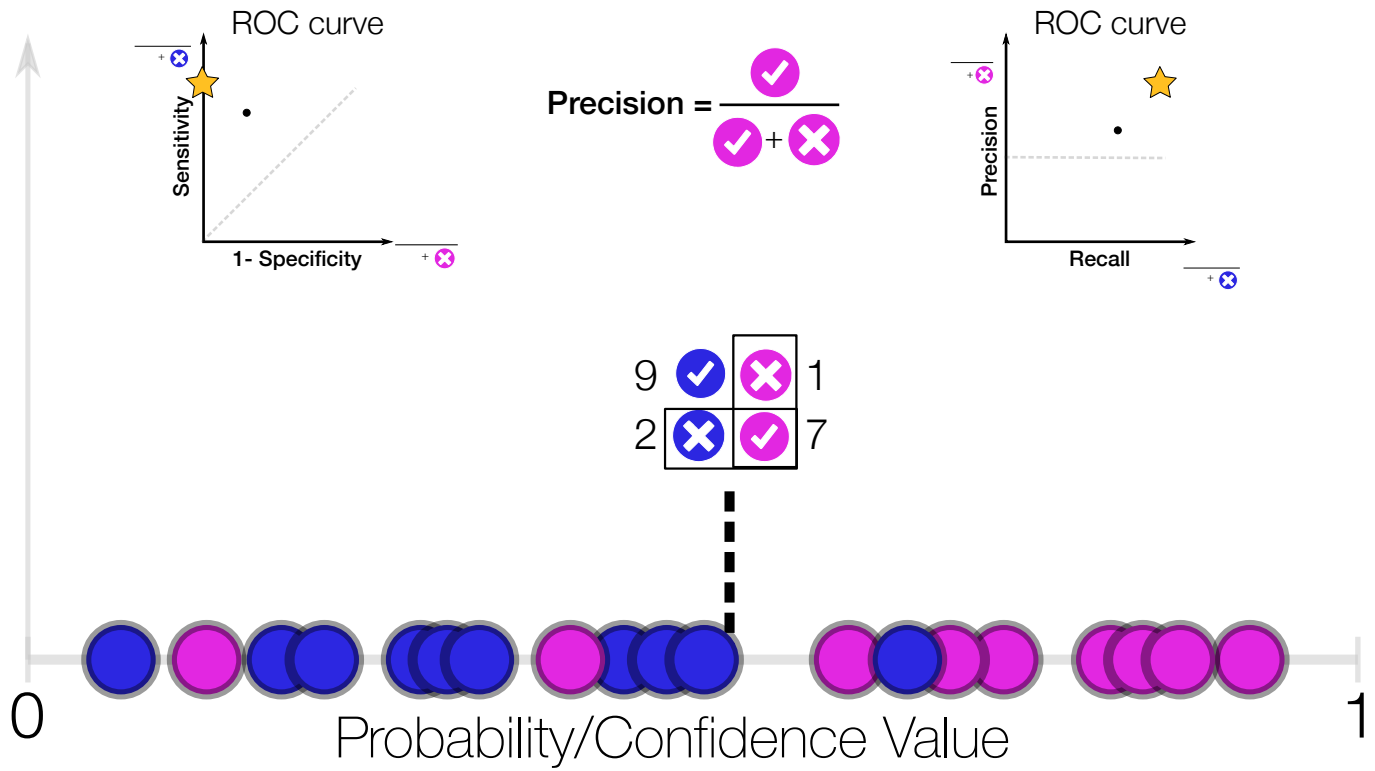
How Good is my Model?



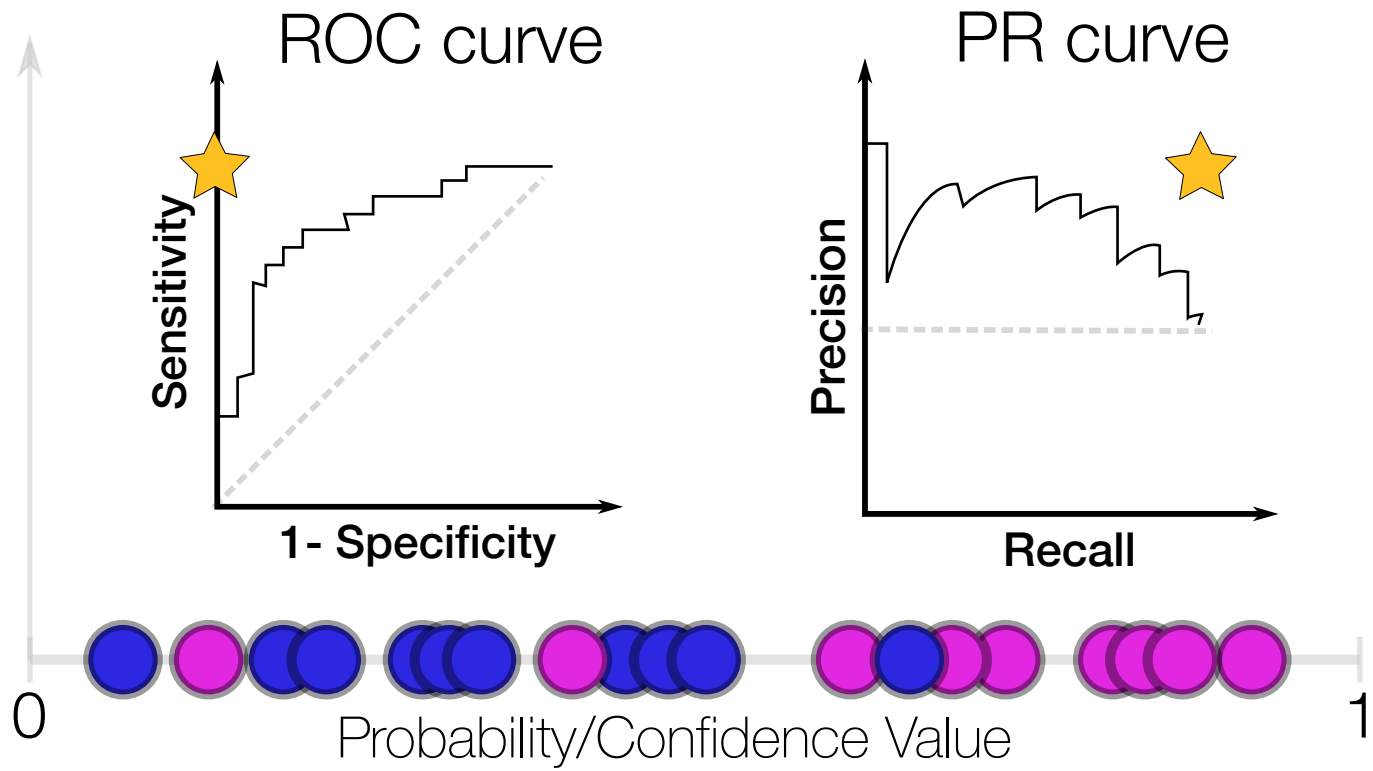
How Good is my Model?



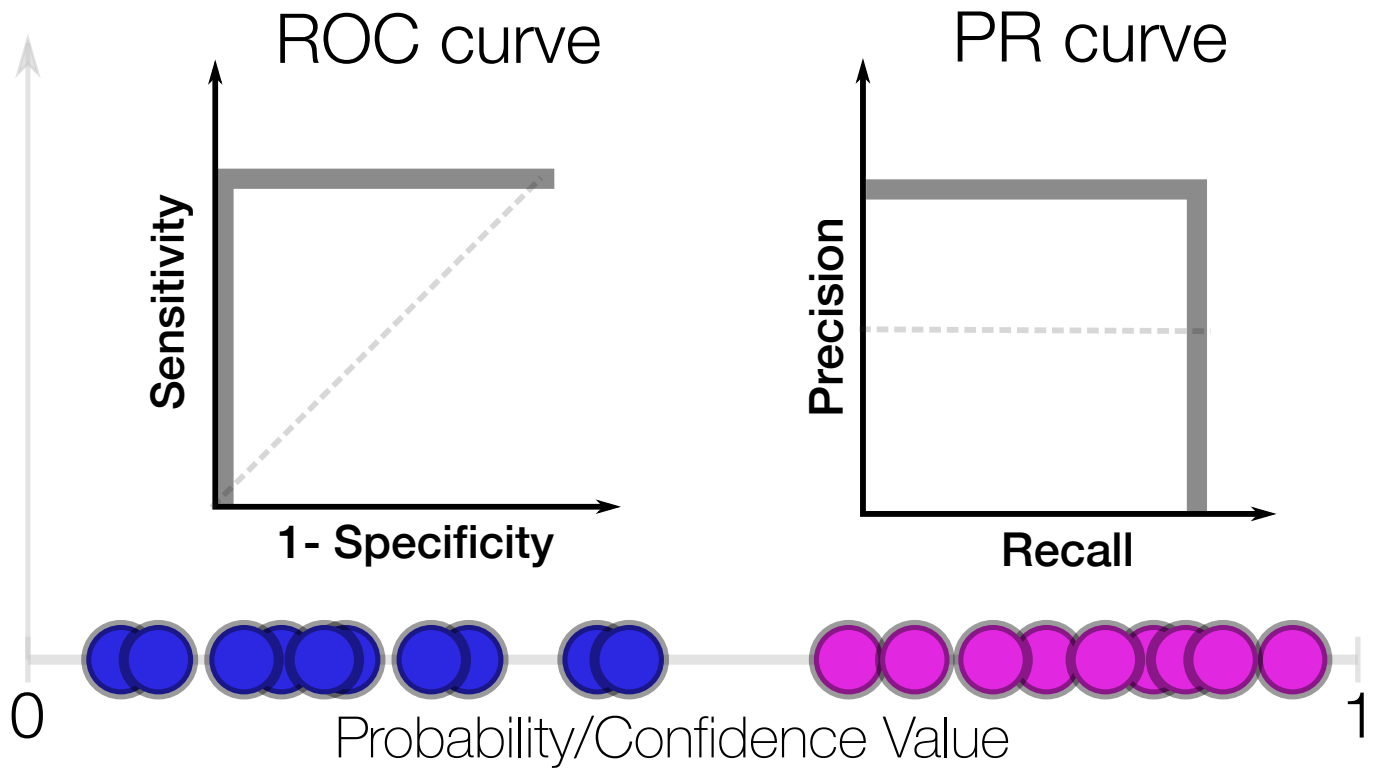
How Good is my Model?



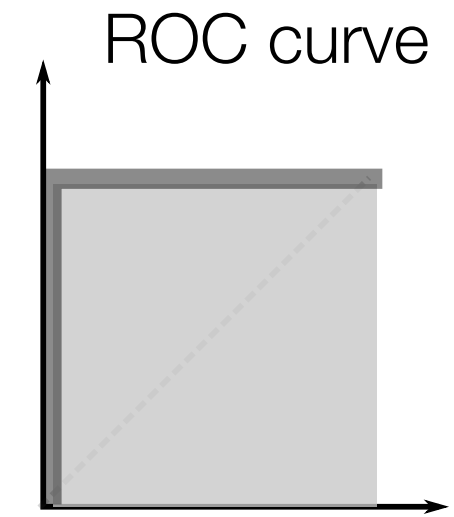
How Good is my Model?



How Good is my Model?

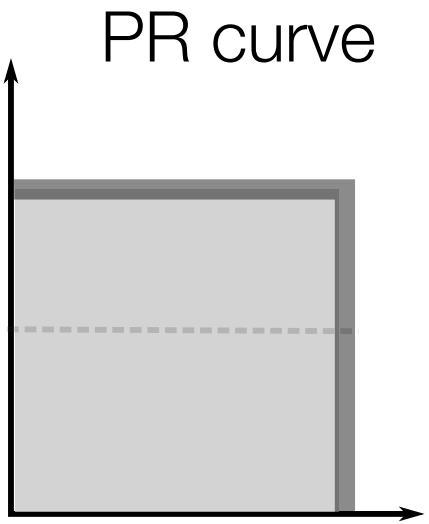


How Good is my Model?



AUC

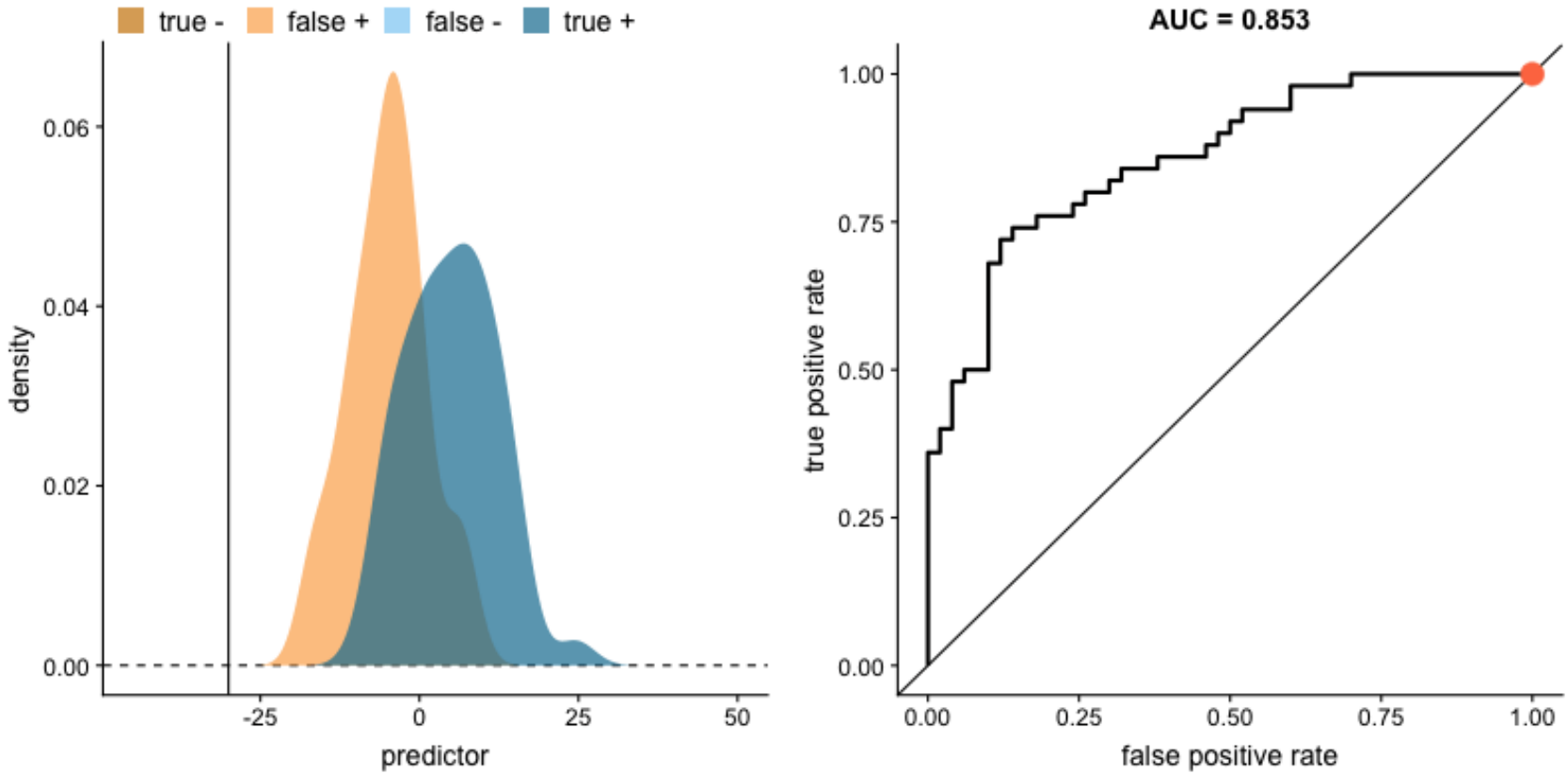
ROC = 1



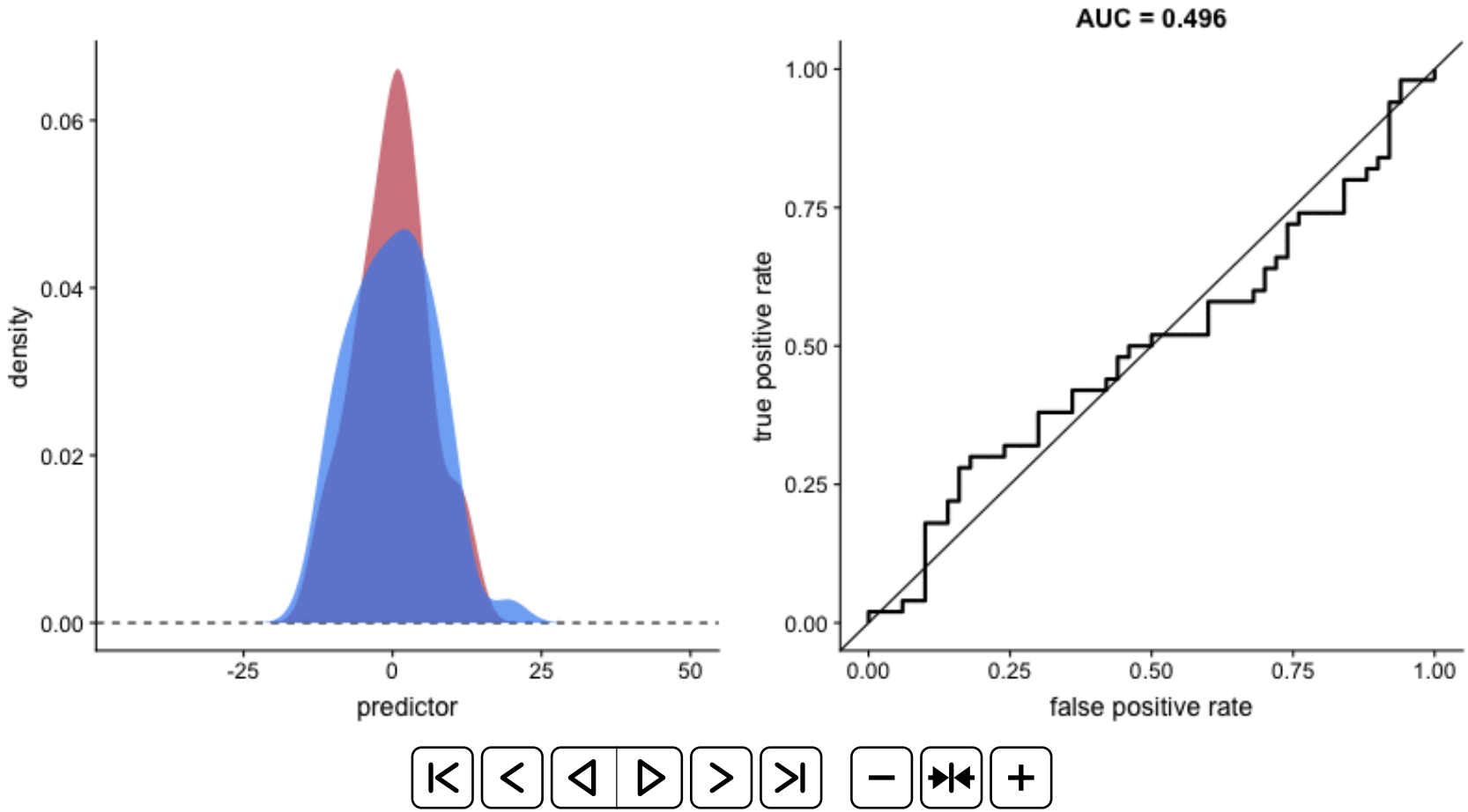
AUC

PR = 1

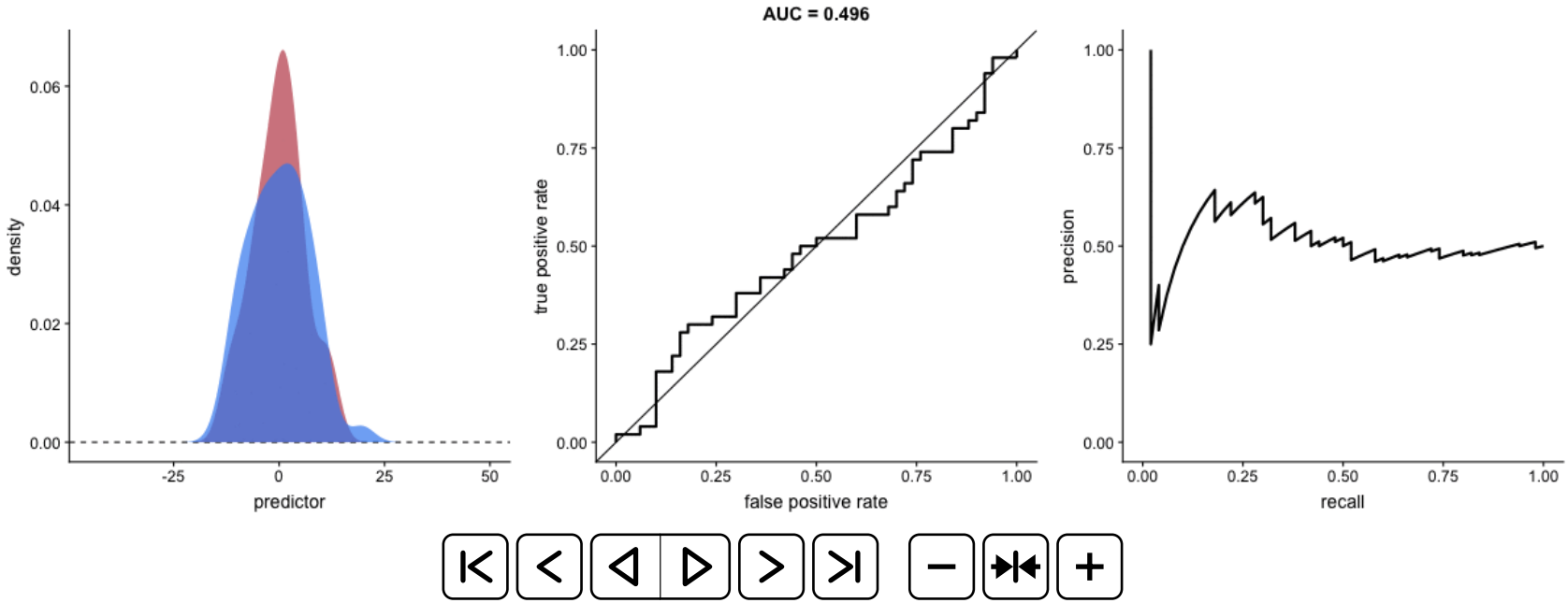
Animation: Variation in Threshold



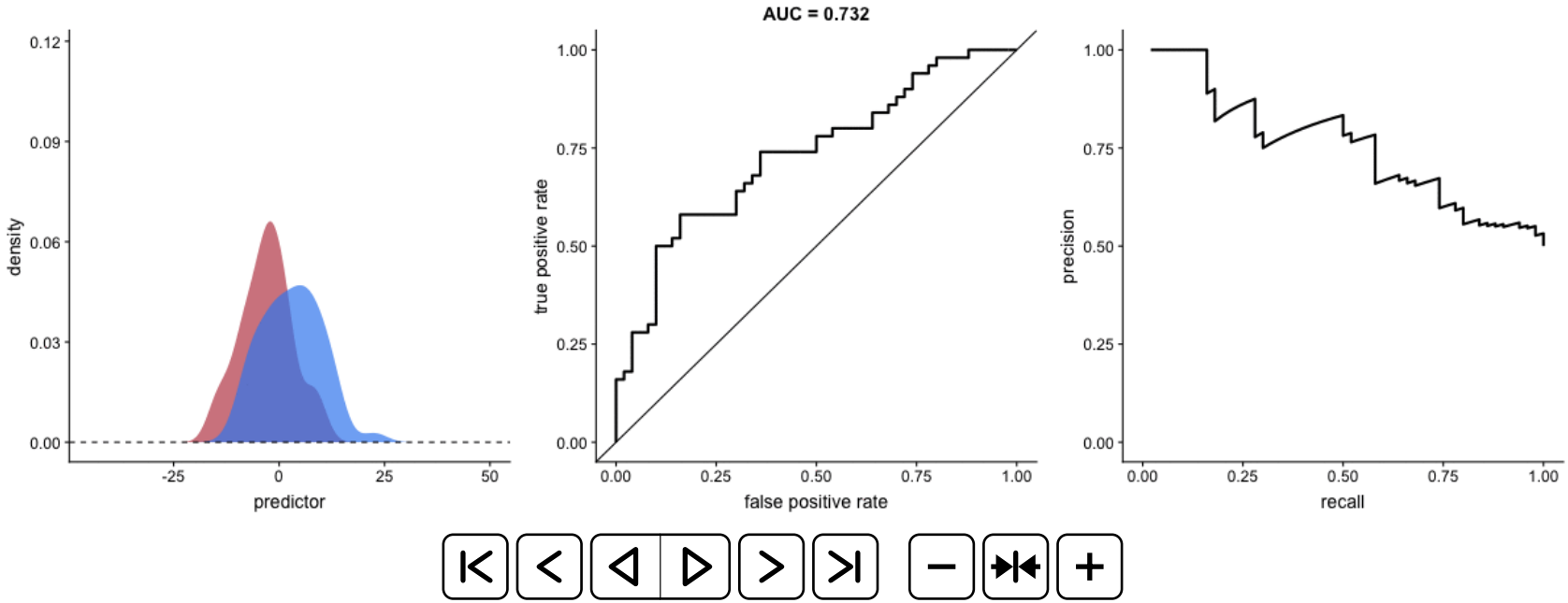
Animation: Improved Models by ROC Curve



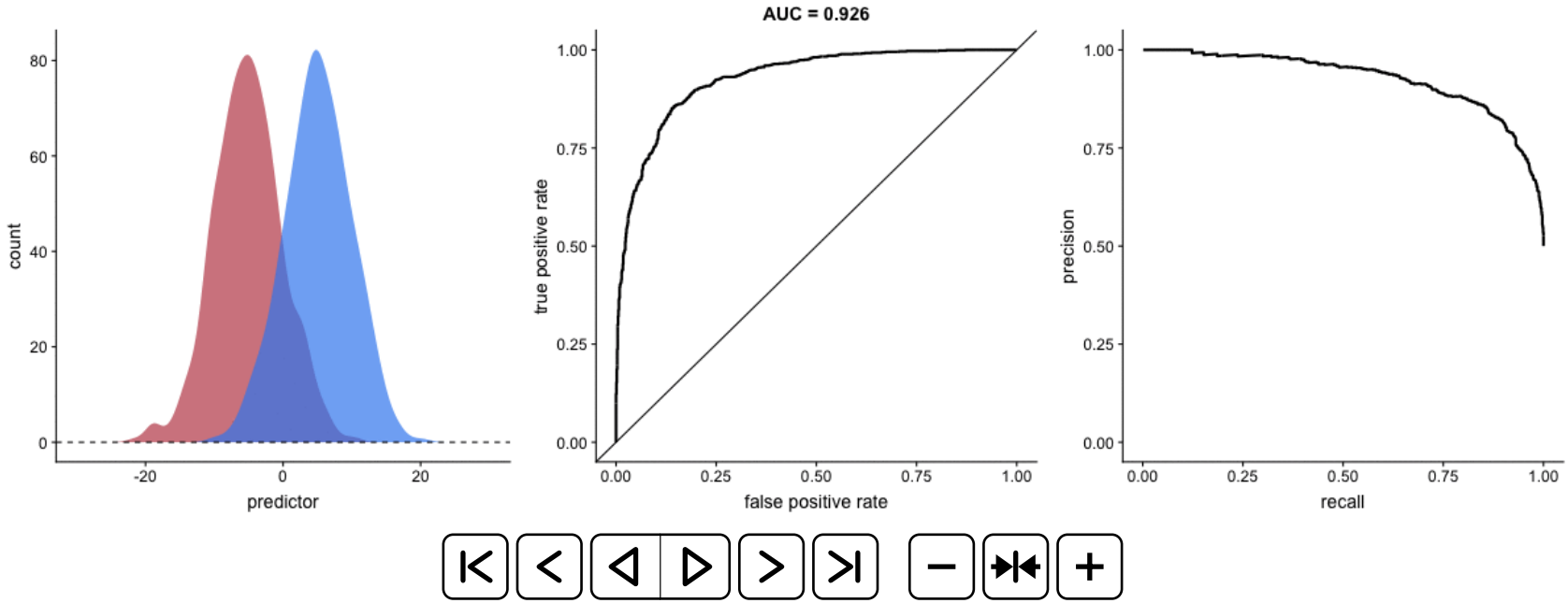
Animation: Improved Models by PR Curve



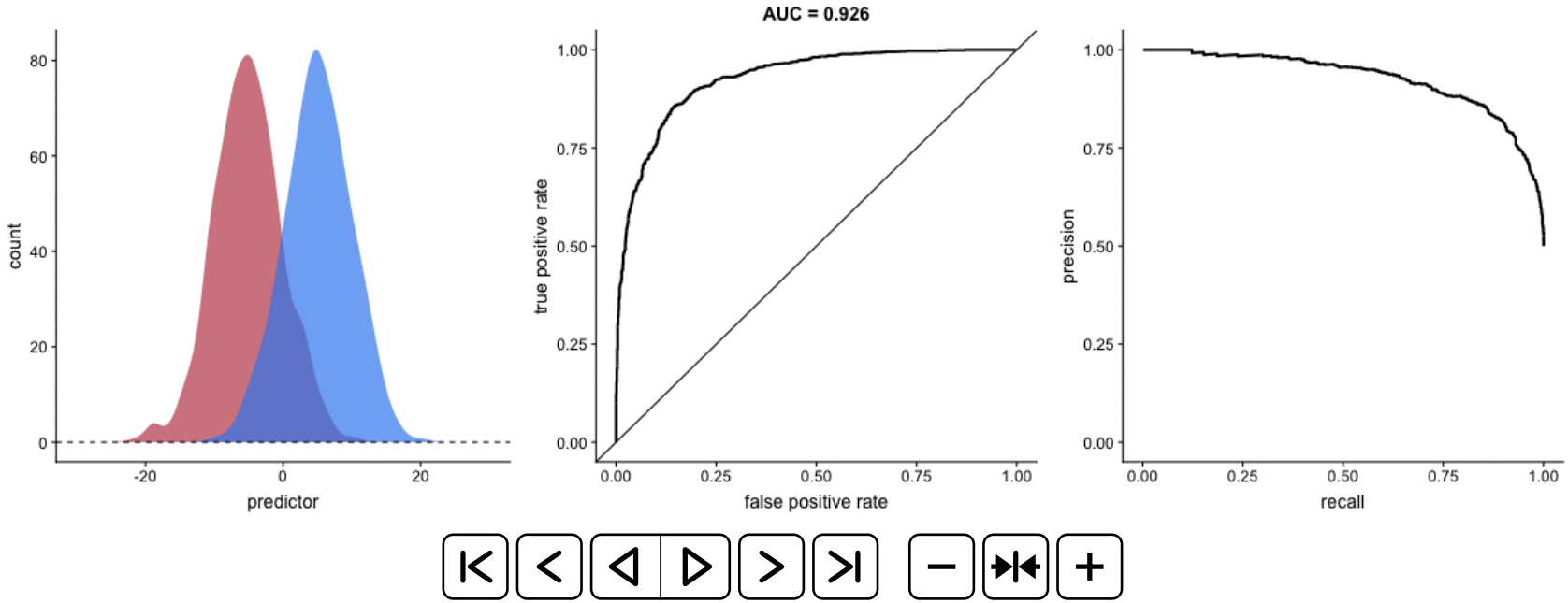
Animation: Variation in Standard Deviation



Animation: Variation in Class Imbalance



Animation: Variation in Class Imbalance



Into the Notebooks we Go...

We will cover one new notebook today!

1. Tutorial 5 - Performance Metrics

Tutorial 5 - Performance Metrics

Kevin Dick, PhDc Biomedical Engineering

Carleton University

Friday 9th October, 2020