Tutorial 3 - DTs, SVMs, KNNs (oh my!)

Kevin Dick, PhDc Biomedical Engineering
Carleton University

Friday 25th September, 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.

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

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 (easy):

- **I** will keep participant video off-screen
- **I** will address questions in the chat by **first name** only
- **You** should stay muted and ask questions in the chat[1].

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

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 (easy):

- **I** will keep participant video off-screen
- **I** will address questions in the chat by **first name** only
- **You** should stay muted and ask questions in the chat[1].

Privacy Preservation (less easy):
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

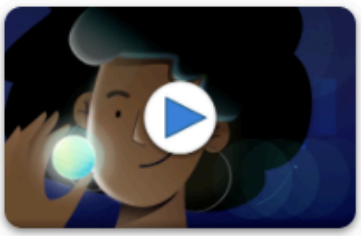
ML Weekly

1. **(NLP)** Go Ahead, Try to Sneak Bad Words Past AI Filters—for Research

Rethinking AI Benchmarking

Dynabench is a research platform for dynamic data collection and benchmarking. Static benchmarks have well-known issues: they saturate quickly, are susceptible to overfitting, contain exploitable annotator artifacts and have unclear or imperfect evaluation metrics.

This platform in essence is a scientific experiment: can we make faster progress if we collect data dynamically, with humans and models in the loop, rather than in the old-fashioned static way?

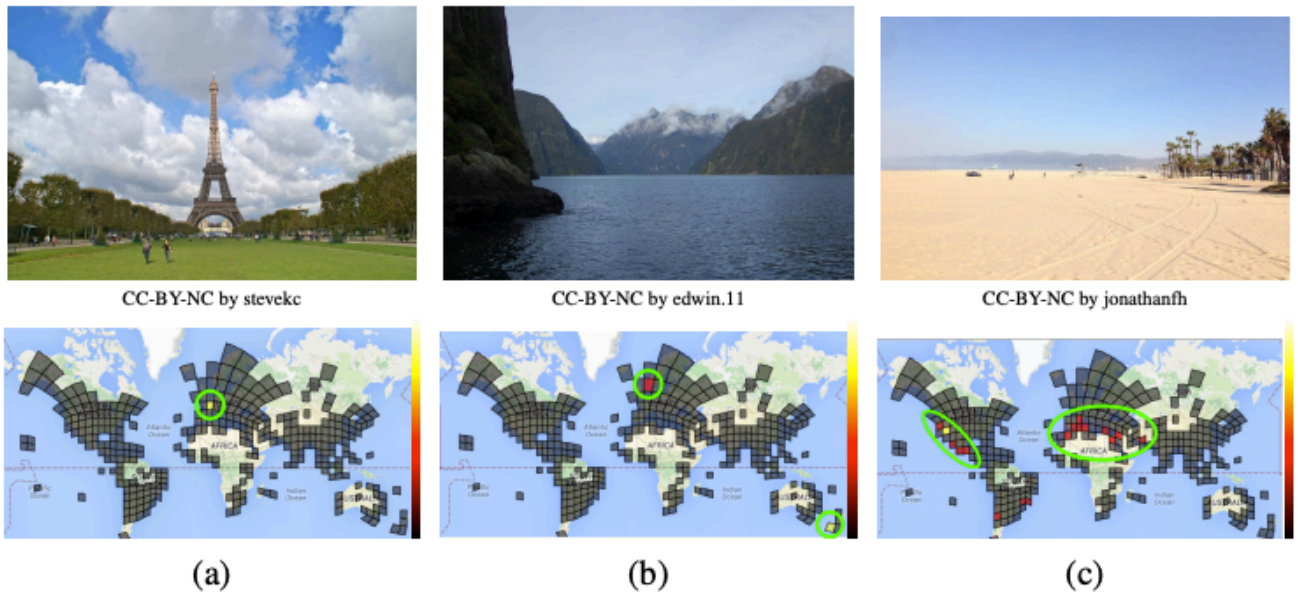


ML Weekly

1. **(NLP)** Go Ahead, Try to Sneak Bad Words Past AI Filters—for Research
2. **(Robotics)** An adaptive deep reinforcement learning framework enables curling robots with human-like performance in real-world conditions

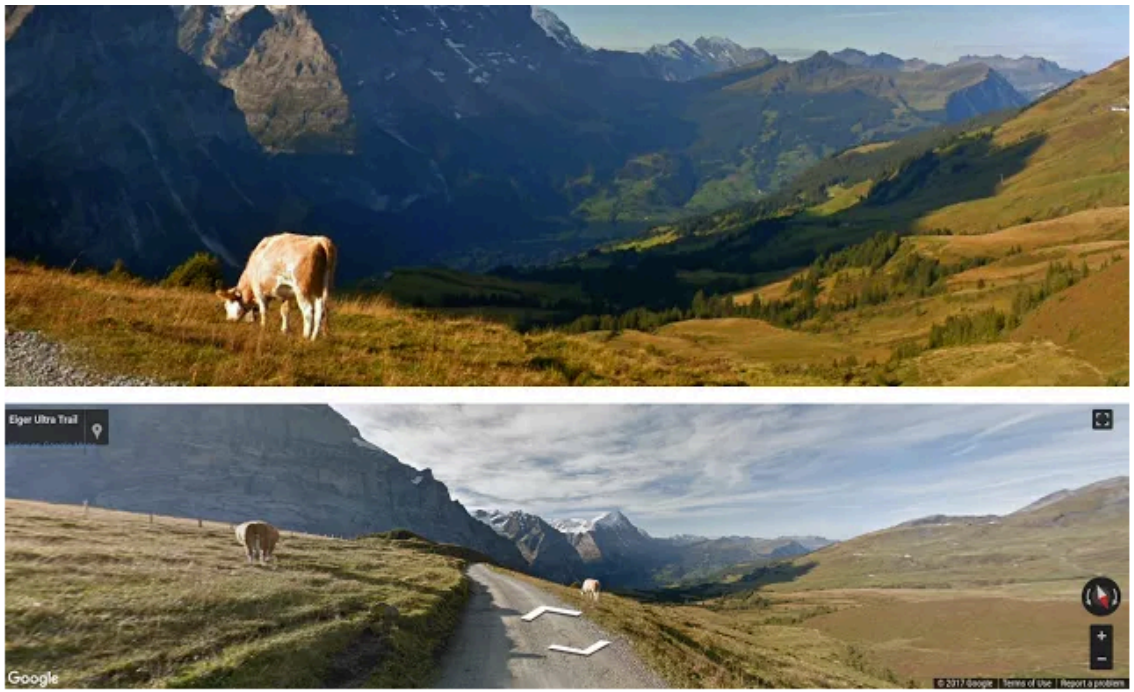
ML Weekly

1. **(NLP)** Go Ahead, Try to Sneak Bad Words Past AI Filters—for Research
2. **(Robotics)** An adaptive deep reinforcement learning framework enables curling robots with human-like performance in real-world conditions
3. **(Vision)** PlaNet: Photo Geolocation with Convolutional Neural Networks



ML Weekly

1. **(NLP)** Go Ahead, Try to Sneak Bad Words Past AI Filters—for Research
2. **(Robotics)** An adaptive deep reinforcement learning framework enables curling robots with human-like performance in real-world conditions
3. **(Vision)** PlaNet: Photo Geolocation with Convolutional Neural Networks
4. **(Vision)** Google is using AI to create stunning landscape photos using Street View imagery



Tutorial Intuition

Building an Intuition for the Concepts of this Tutorial

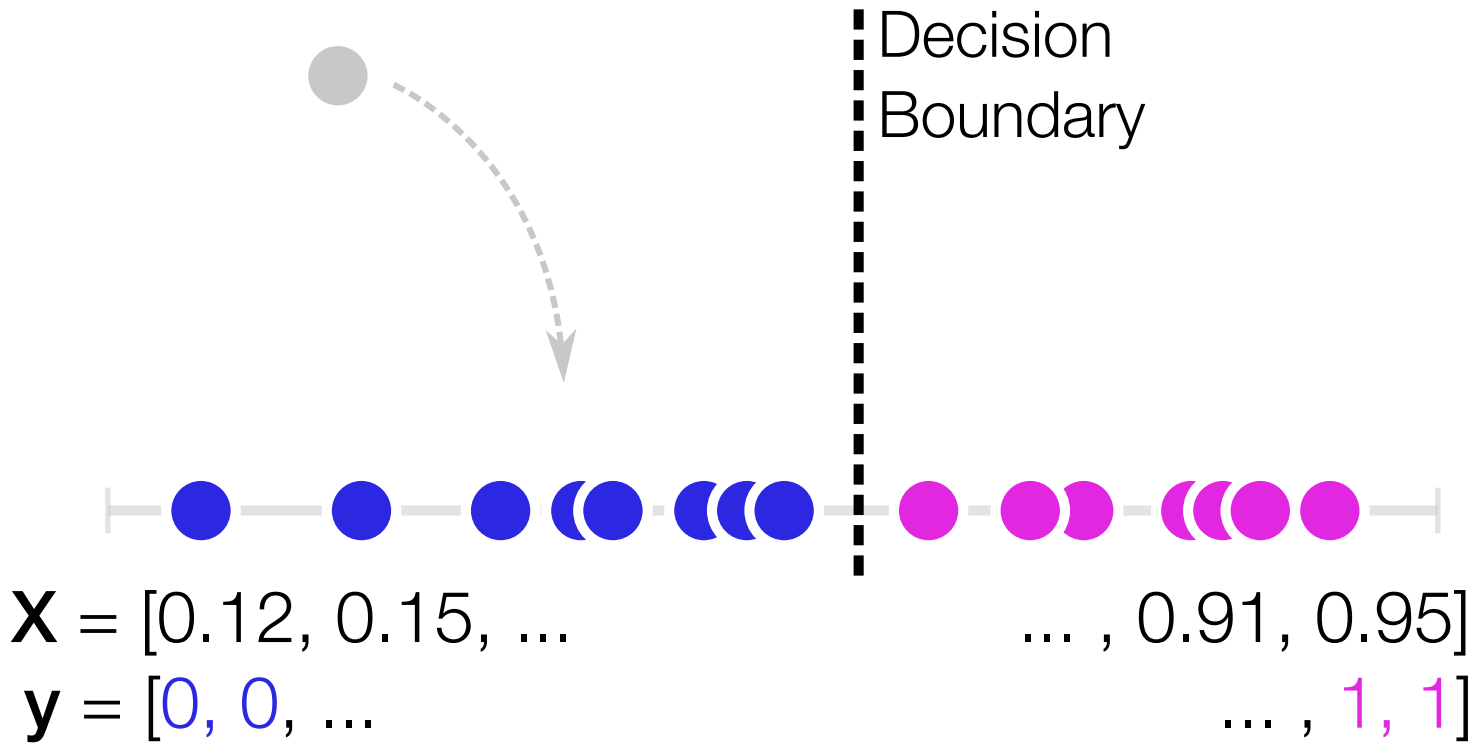
Creating and Visualizing Decision Boundaries

Creating and Visualizing Decision Boundaries

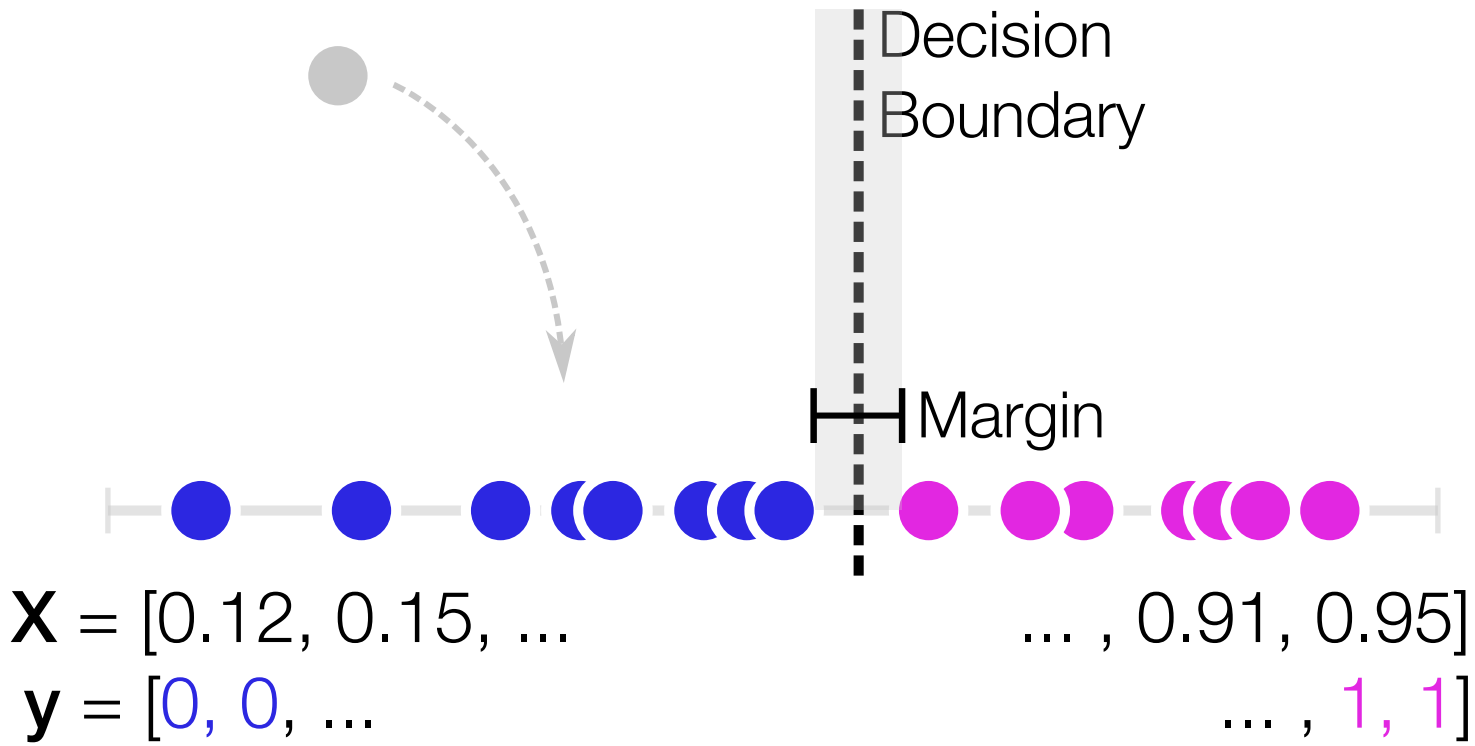
$\mathbf{X}$ = [0.12, 0.15, ... , 0.91, 0.95]

$\mathbf{y}$ = [0, 0, ... , 1, 1]

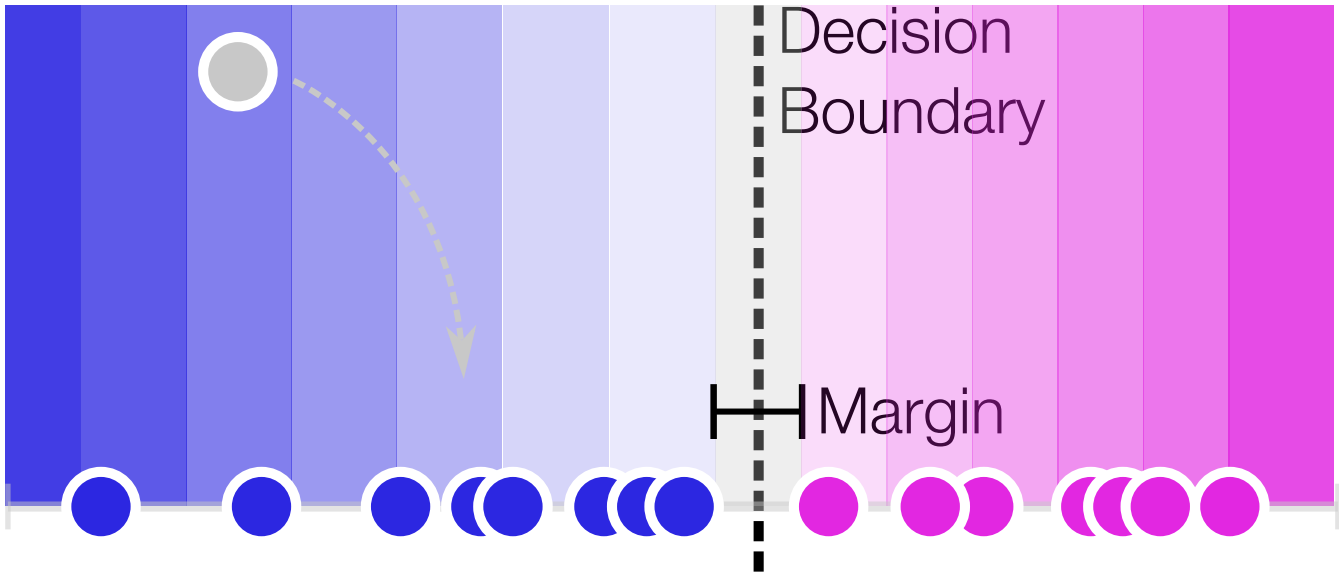
Creating and Visualizing Decision Boundaries



Creating and Visualizing Decision Boundaries



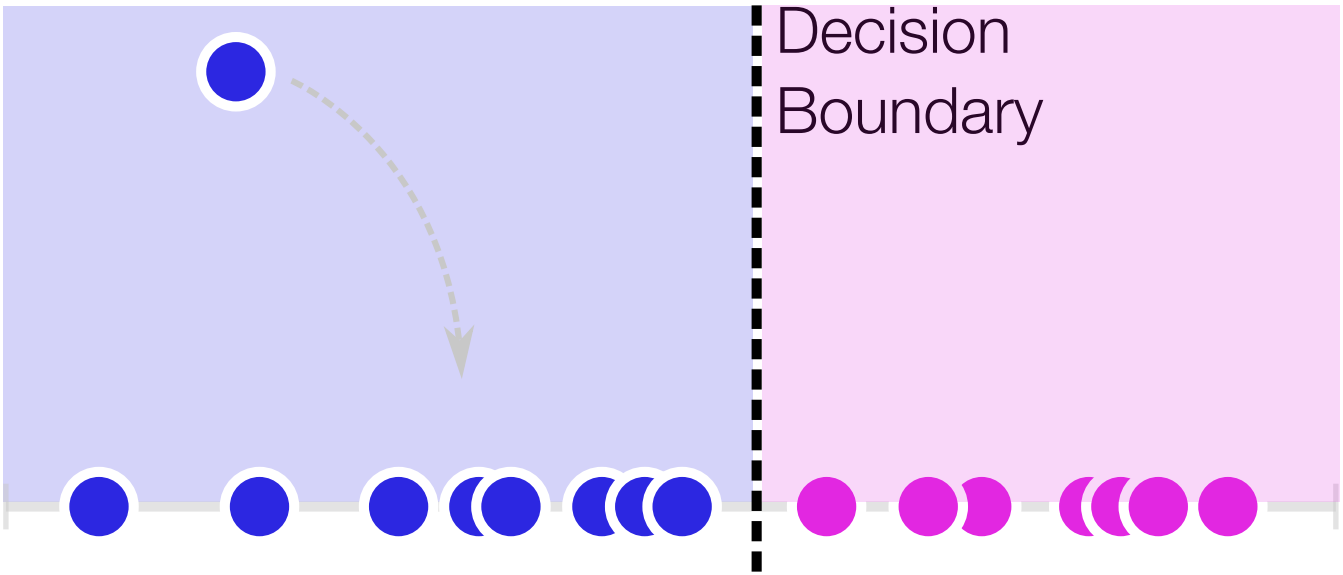
Creating and Visualizing Decision Boundaries



X = [0.12, 0.15, ... , 0.91, 0.95]

y = [0, 0, ... , 1, 1]

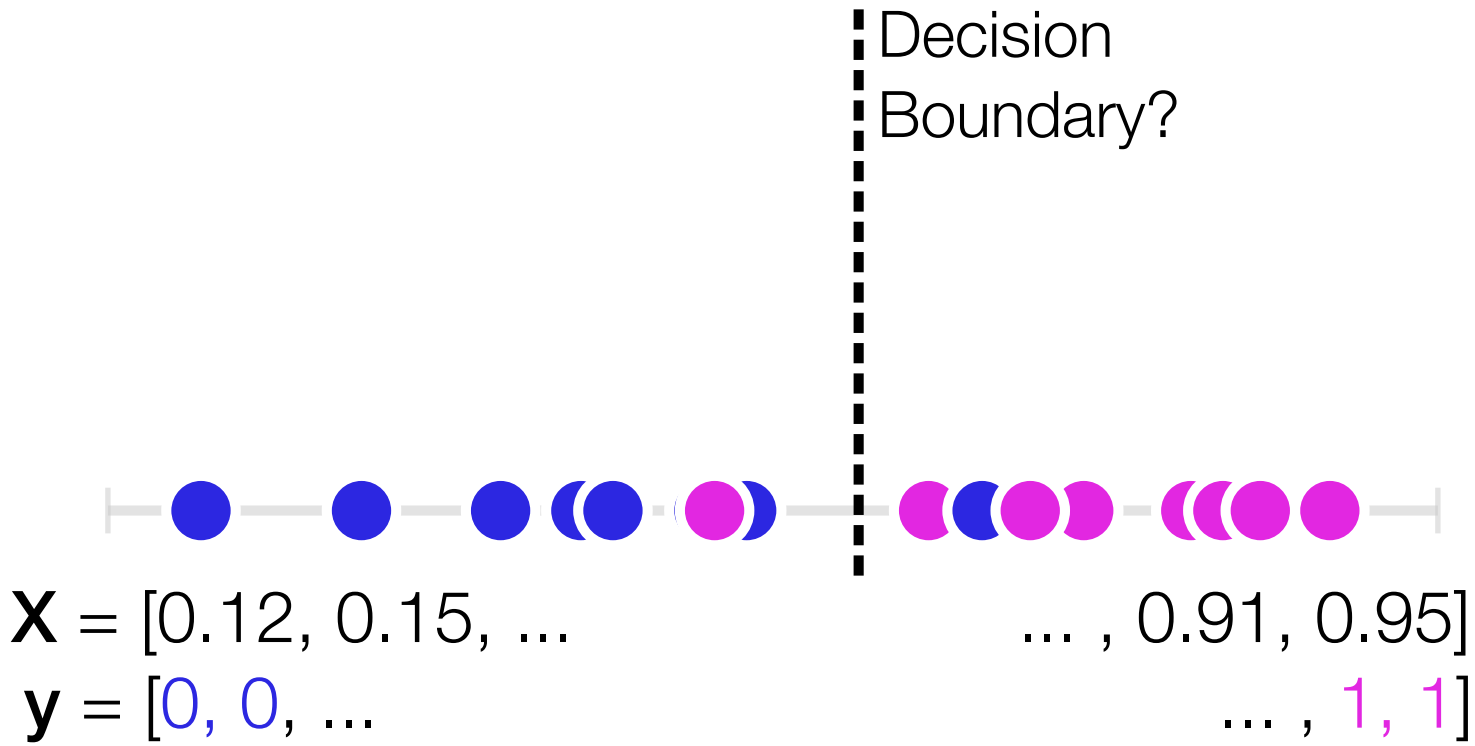
Creating and Visualizing Decision Boundaries



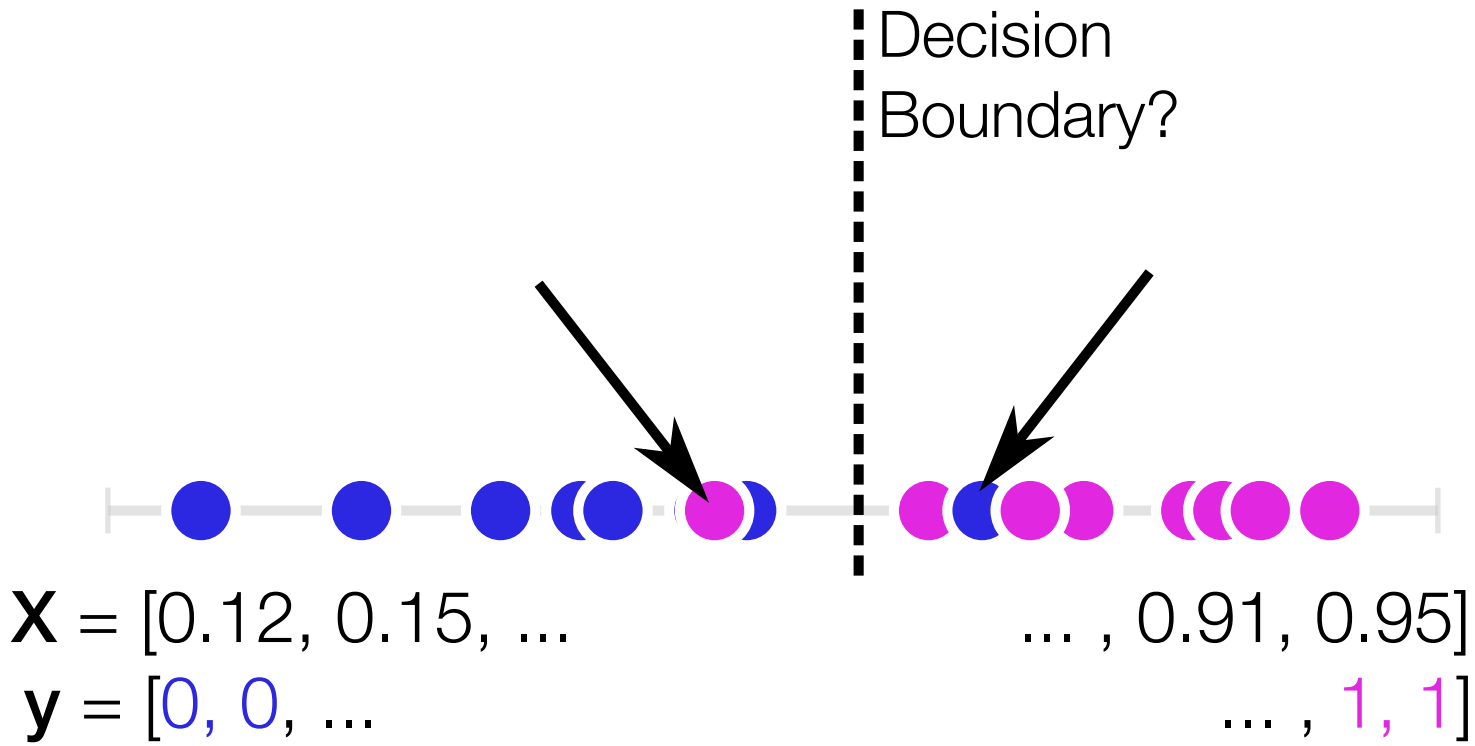
X = [0.12, 0.15, ... , 0.91, 0.95]

y = [0, 0, ... , 1, 1]

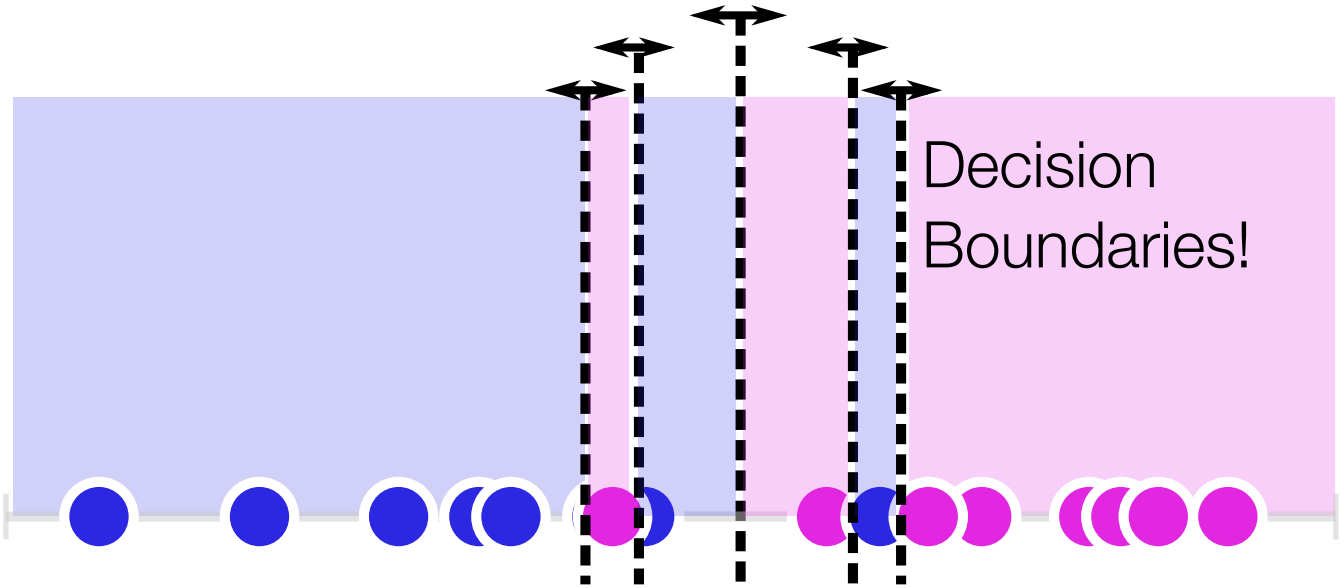
Creating and Visualizing Decision Boundaries



Creating and Visualizing Decision Boundaries



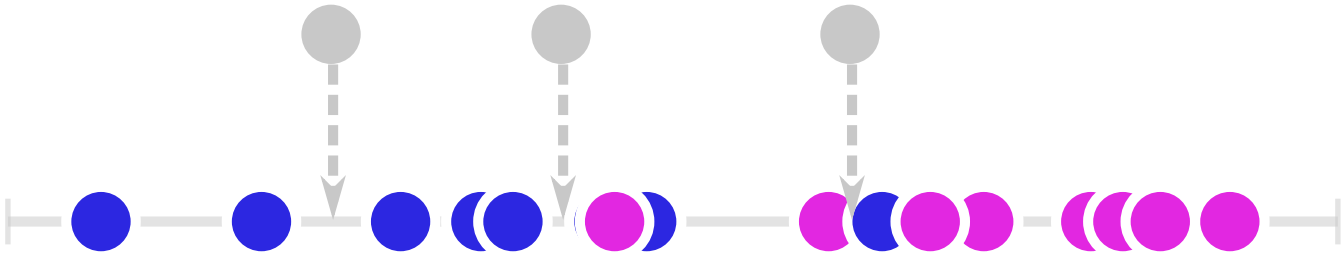
Creating and Visualizing Decision Boundaries



X = [0.12, 0.15, ... , 0.91, 0.95]

y = [0, 0, ... , 1, 1]

Creating and Visualizing Decision Boundaries



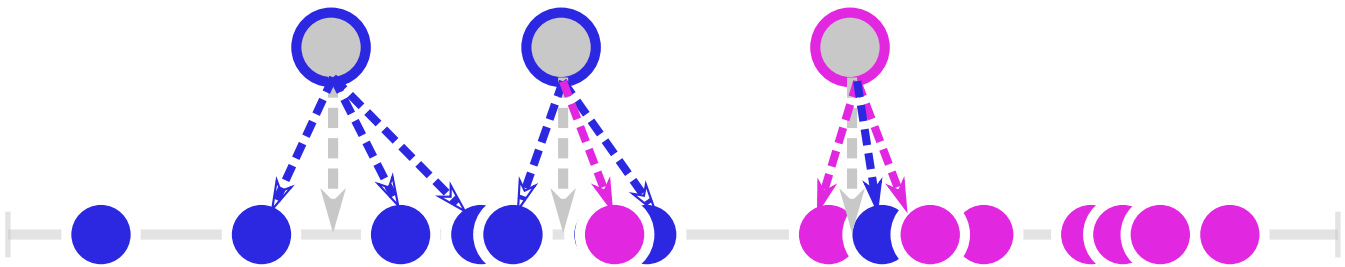
$\mathbf{X}$ = [0.12, 0.15, ... , 0.91, 0.95]

$\mathbf{y}$ = [0, 0, ... , 1, 1]

Creating and Visualizing Decision Boundaries

Instance-Based Method

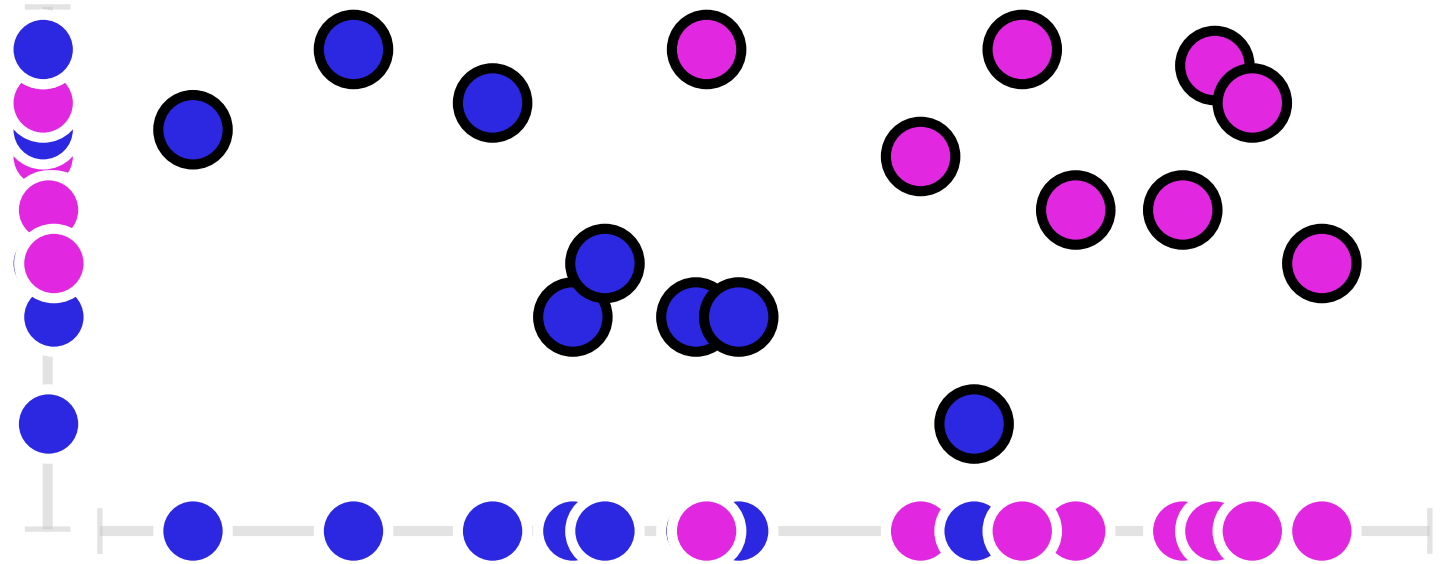
3-Nearest Neighbours



X = [0.12, 0.15, ... , 0.91, 0.95]

y = [0, 0, ... , 1, 1]

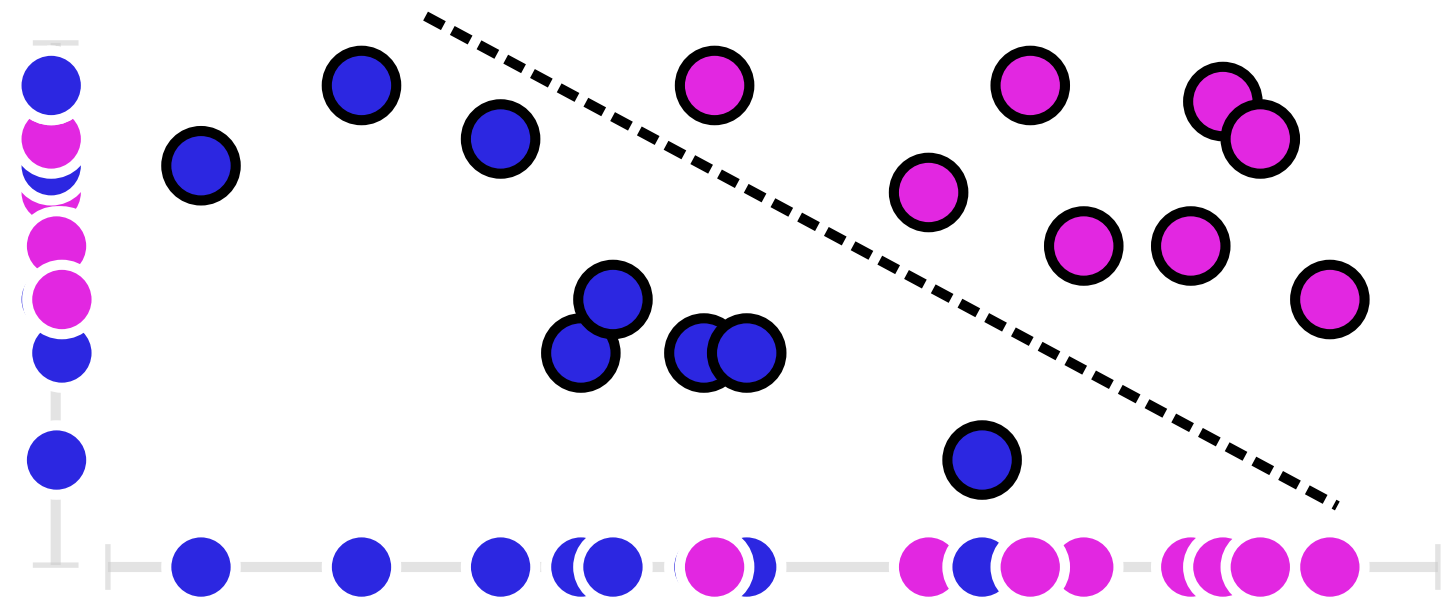
Creating and Visualizing Decision Boundaries



$$\mathbf{X} = \begin{bmatrix} 0.12, 0.15, \ldots & \ldots, 0.91, 0.95 \\ 0.02, 0.09, \ldots & \ldots, 0.93, 0.97 \end{bmatrix}$$

$$\mathbf{y} = [0, 0, \ldots \quad \ldots, 1, 1]$$

Creating and Visualizing Decision Boundaries



$$\mathbf{X} = \begin{bmatrix} 0.12, 0.15, \ldots & \ldots, 0.91, 0.95 \\ 0.02, 0.09, \ldots & \ldots, 0.93, 0.97 \end{bmatrix}$$

$$\mathbf{y} = [0, 0, \ldots \quad \ldots, 1, 1]$$

Comparing "Classical" ML Algorithms

In this tutorial, we will build and compare *classical* learning algorithms and demonstrate where each are more or less useful.

Key Takeaway: no one method is a *Silver Bullet*!

Notebooks

We will cover one new notebook and (with enough time) cover a previous tutorial.

1. Tutorial 3 - DTs, SVMs, KNNs (oh my!)
2. Tutorial 2 - The Prototypical Machine Learning Notebook (Iris Dataset)

Tutorial 3 - DTs, SVMs, KNNs (oh my!)

Kevin Dick, PhDc Biomedical Engineering
Carleton University

Friday 25th September, 2020