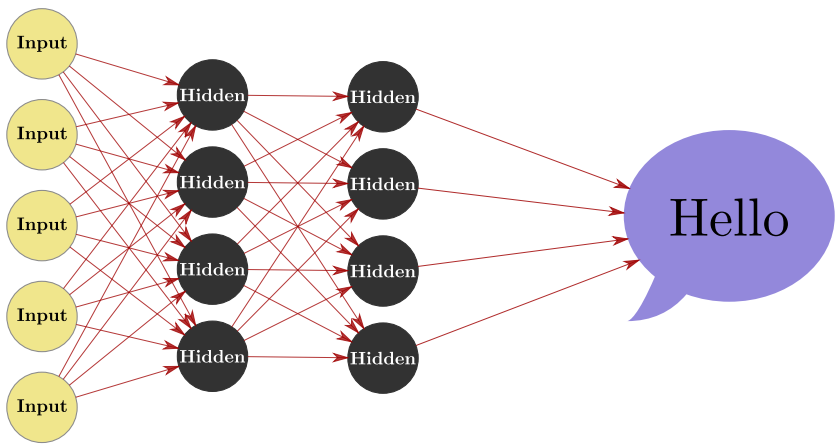
Introduction to Machine Learning

Pavlos Vougiouklis



The Plan (hopefully?)

- Short overview talk on classical Machine Learning
- Practical session using Jupyter Notebook (Python 2), scikit-learn and pandas to experiment with feature extraction, classification and clustering
- Please ask questions as we go :)

Why Should We Care?

- ML allows us to directly learn from data without requiring hand-coded rules
- The ever-increasing amount of training data on the Web has benefited the performance ML algorithms substantially
- Machine Learning models are matrix-oriented models; advances in Parallel Computing along with the sophistication of the tools for GPU Computing have substantially increased their computing capabilities
- ML has achieved super-human performance in domains such as visual object recognition or (video) games (e.g. Go, chess or Dota 2)

Why Should We Care?

- ML allows us to directly learn from data without requiring hand-coded rules
- The ever-increasing amount of training data on the Web has benefited the performance ML algorithms substantially
- Machine Learning models are matrix-oriented models; advances in Parallel Computing along with the sophistication of the tools for GPU Computing have substantially increased their computing capabilities
- ML has achieved super-human performance in domains such as visual object recognition or (video) games (e.g. Go, chess or Dota 2)

Why Should We Care?

- ML allows us to directly learn from data without requiring hand-coded rules
- The ever-increasing amount of training data on the Web has benefited the performance ML algorithms substantially
- Machine Learning models are matrix-oriented models; advances in Parallel Computing along with the sophistication of the tools for GPU Computing have substantially increased their computing capabilities
- ML has achieved super-human performance in domains such as visual object recognition or (video) games (e.g. Go, chess or Dota 2)

Why Should We Care?

- ML allows us to directly learn from data without requiring hand-coded rules
- The ever-increasing amount of training data on the Web has benefited the performance ML algorithms substantially
- Machine Learning models are matrix-oriented models; advances in Parallel Computing along with the sophistication of the tools for GPU Computing have substantially increased their computing capabilities
- ML has achieved super-human performance in domains such as visual object recognition or (video) games (e.g. Go, chess or Dota 2)

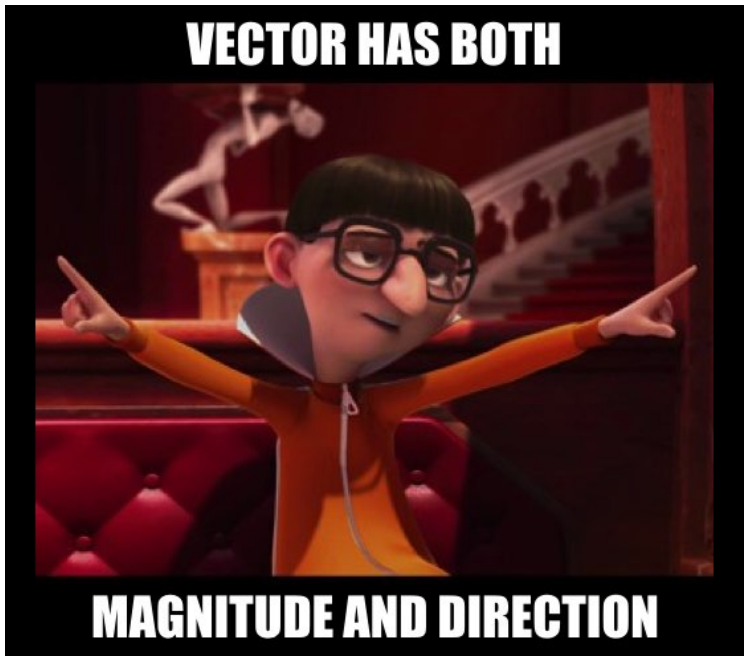
Why Should We Care?

- ML allows us to directly learn from data without requiring hand-coded rules
- The ever-increasing amount of training data on the Web has benefited the performance ML algorithms substantially
- Machine Learning models are matrix-oriented models; advances in Parallel Computing along with the sophistication of the tools for GPU Computing have substantially increased their computing capabilities
- ML has achieved super-human performance in domains such as visual object recognition or (video) games (e.g. Go, chess or Dota 2)

Why Should We Care?

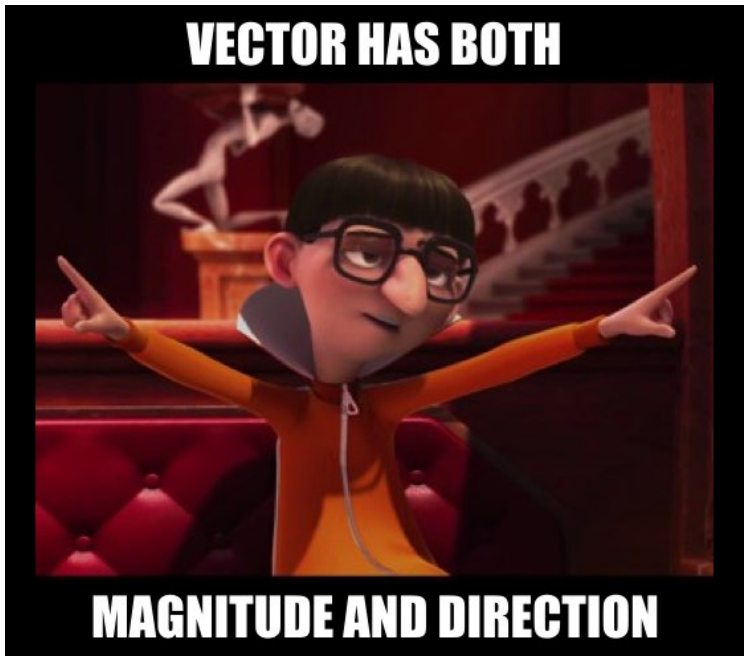
- ML allows us to directly learn from data without requiring hand-coded rules
- The ever-increasing amount of training data on the Web has benefited the performance ML algorithms substantially
- Machine Learning models are matrix-oriented models; advances in Parallel Computing along with the sophistication of the tools for GPU Computing have substantially increased their computing capabilities
- ML has achieved super-human performance in domains such as visual object recognition or (video) games (e.g. Go, chess or Dota 2); in others, such as Natural Language Understanding or dialogue systems, it's still far from perfect

The Foundation of Machine Learning



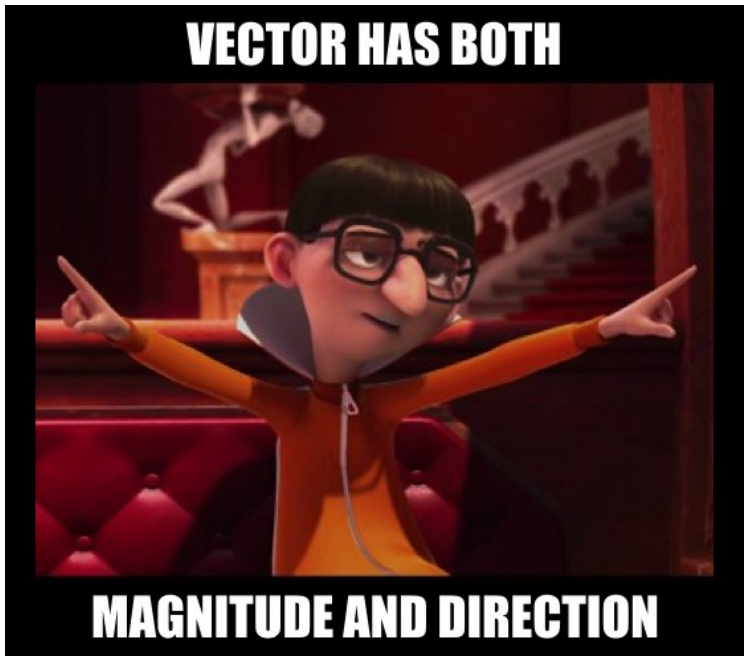
- Machine Learning is a **highly mathematical** subject

The Foundation of Machine Learning



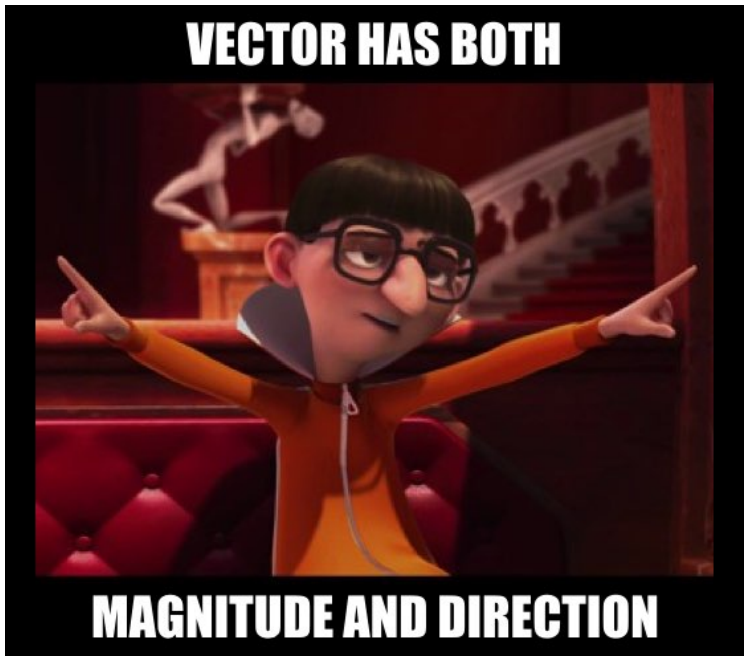
- Machine Learning is a **highly mathematical** subject

The Foundation of Machine Learning



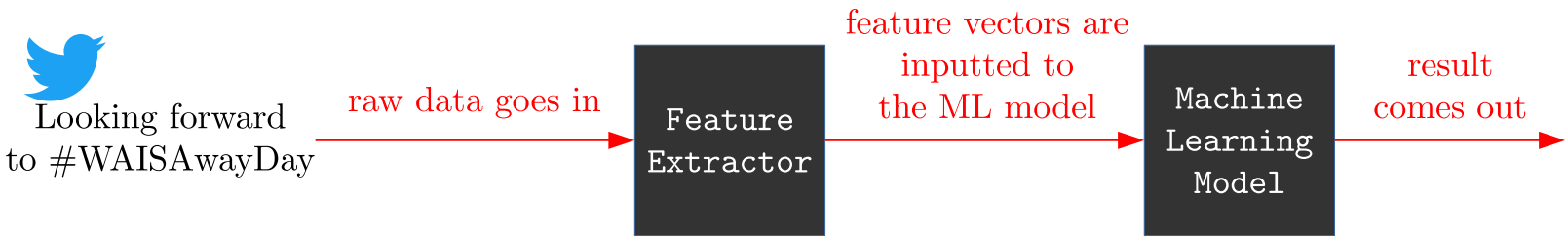
- Machine Learning is a **highly mathematical** subject; Today, we will try to ignore the details
- The **vector** is the key concept in Machine Learning

The Foundation of Machine Learning



- Machine Learning is a **highly mathematical** subject; Today, we will try to ignore the details
- The **vector** is the key concept in Machine Learning

The General Machine Learning Pipeline



Feature Extraction

- It is the process of transforming raw data into feature-vectors, which are mathematical vectors
  - ⋆ list of (usually) Real numbers
  - ⋆ fixed number of elements; the number of element is the dimensionality of the vector
- Each feature-vector represents a point in a feature-space or equally a direction in the feature-space
- The dimensionality of a feature-space is the dimensionality of every vector within it; vectors of different dimensionality cannot exist in the same feature-space

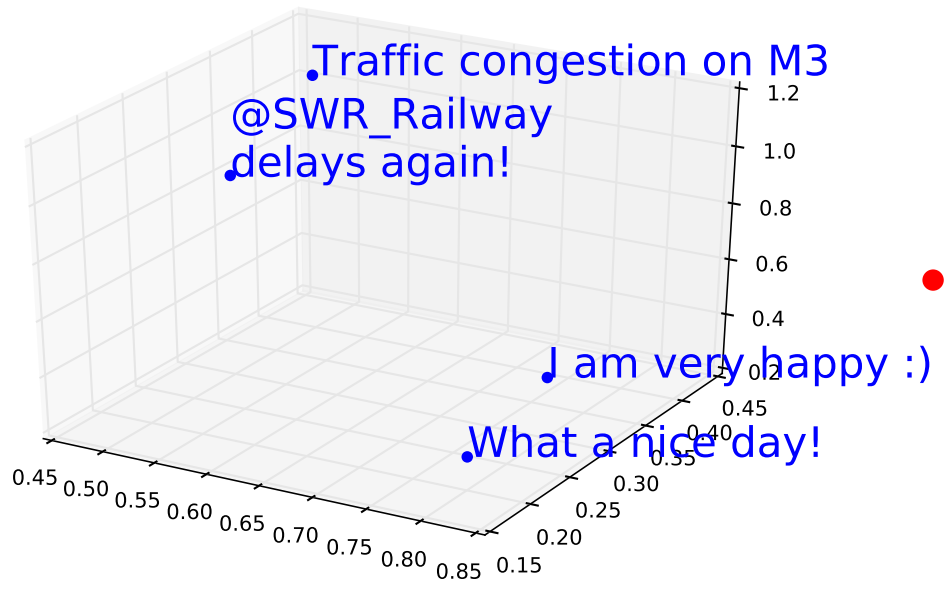
Feature Extraction

- It is the process of transforming raw data into feature-vectors, which are mathematical vectors
  - ⋆ list of (usually) Real numbers
  - ⋆ fixed number of elements; the number of element is the dimensionality of the vector
- Each feature-vector represents a point in a feature-space or equally a direction in the feature-space
- The dimensionality of a feature-space is the dimensionality of every vector within it; vectors of different dimensionality cannot exist in the same feature-space

Feature Extraction

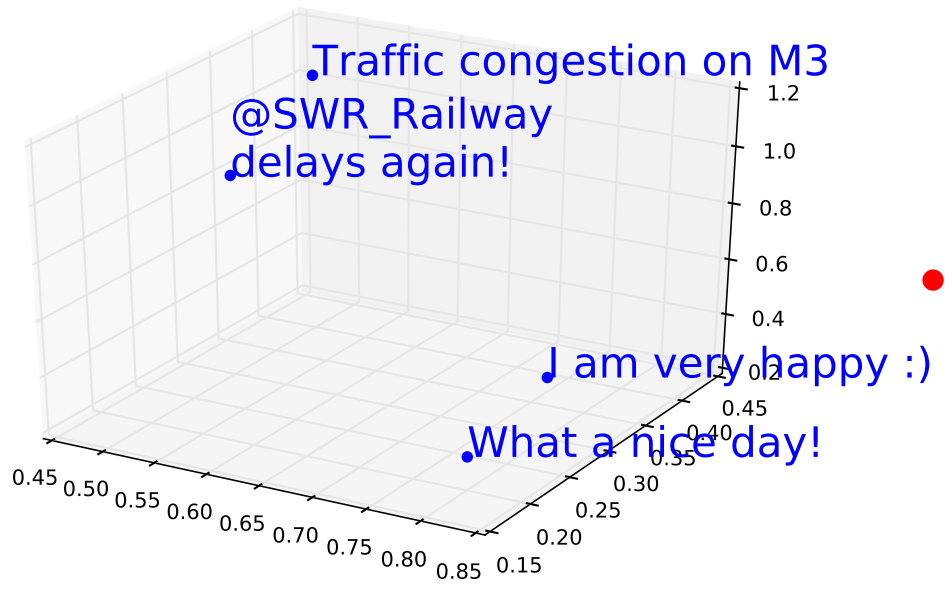
- It is the process of transforming raw data into feature-vectors, which are mathematical vectors
  - ⋆ list of (usually) Real numbers
  - ⋆ fixed number of elements; the number of element is the dimensionality of the vector
- Each feature-vector represents a point in a feature-space or equally a direction in the feature-space
- The dimensionality of a feature-space is the dimensionality of every vector within it; vectors of different dimensionality cannot exist in the same feature-space

Distances In Feature-Space



- The general goal of a feature extractor is to produce feature-vectors that are close together for *similar* inputs
- Closeness of two vectors in a feature-space can be computed by measuring the distance between the vectors

Distances In Feature-Space



- The general goal of a feature extractor is to produce feature-vectors that are close together for *similar* inputs
- Closeness of two vectors in a feature-space can be computed by measuring the distance between the vectors

How To Choose Features?

- Choose features that allow you to distinguish objects or classes of interest
  - ⋆ similar within classes
  - ⋆ different between classes
- Try to keep the total number of features small; Machine Learning becomes harder as the dimensionality of the feature-space increases

How To Choose Features?

- Choose features that allow you to distinguish objects or classes of interest
  - ⋆ similar within classes
  - ⋆ different between classes
- Try to keep the total number of features small; Machine Learning becomes harder as the dimensionality of the feature-space increases

Supervised Machine Learning: Classification

- Classification is the process of assigning a **class label** to an item (e.g. an image or a piece of text)
- A supervised machine learning algorithm uses a set of pre-labelled *training data* to learn how to assign labels to feature-vectors and their corresponding items
- Binary classification is when a classifier has two classes, and multi-class when it has many

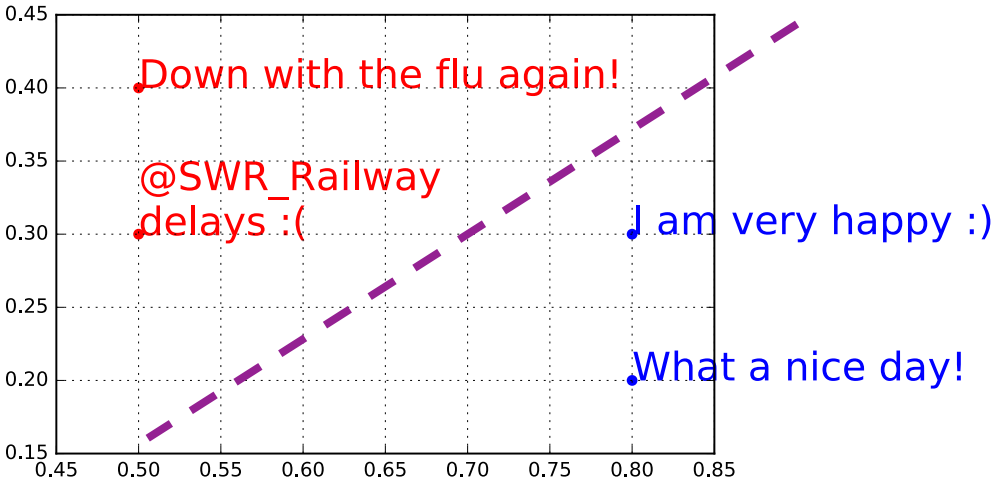
Supervised Machine Learning: Classification

- Classification is the process of assigning a **class label** to an item (e.g. an image or a piece of text)
- A supervised machine learning algorithm uses a set of pre-labelled *training data* to learn how to assign labels to feature-vectors and their corresponding items
- Binary classification is when a classifier has two classes, and multi-class when it has many

Supervised Machine Learning: Classification

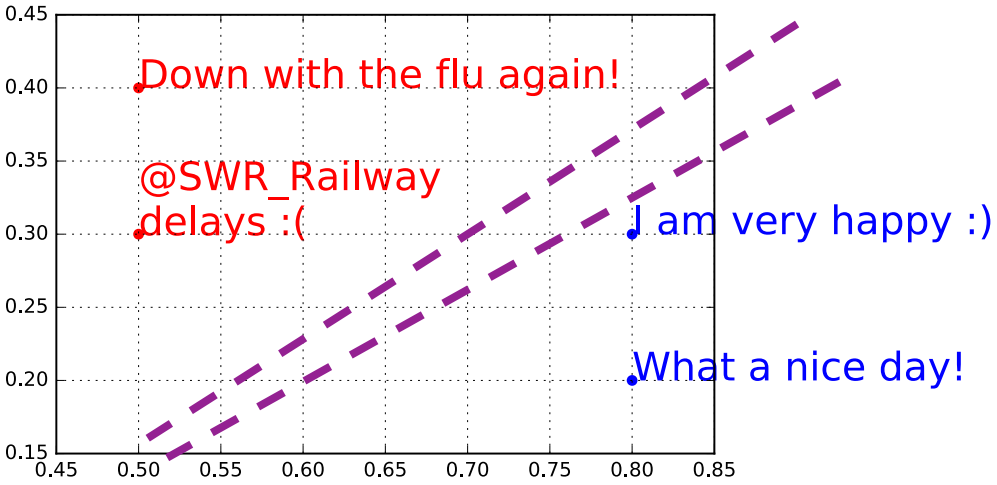
- Classification is the process of assigning a **class label** to an item (e.g. an image or a piece of text)
- A supervised machine learning algorithm uses a set of pre-labelled *training data* to learn how to assign labels to feature-vectors and their corresponding items
- Binary classification is when a classifier has two classes, and multi-class when it has many

Linear Classifiers



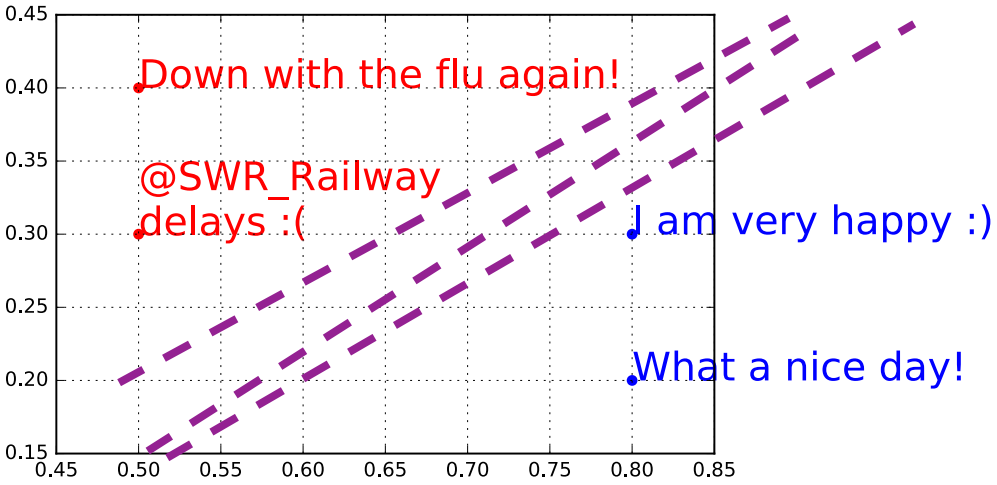
- Linear classifiers try to learn a hyperplane that separates two classes in feature-space with **minimum error**
- Lots of hyperplanes to choose from since different linear classifiers apply different constraints during training

Linear Classifiers



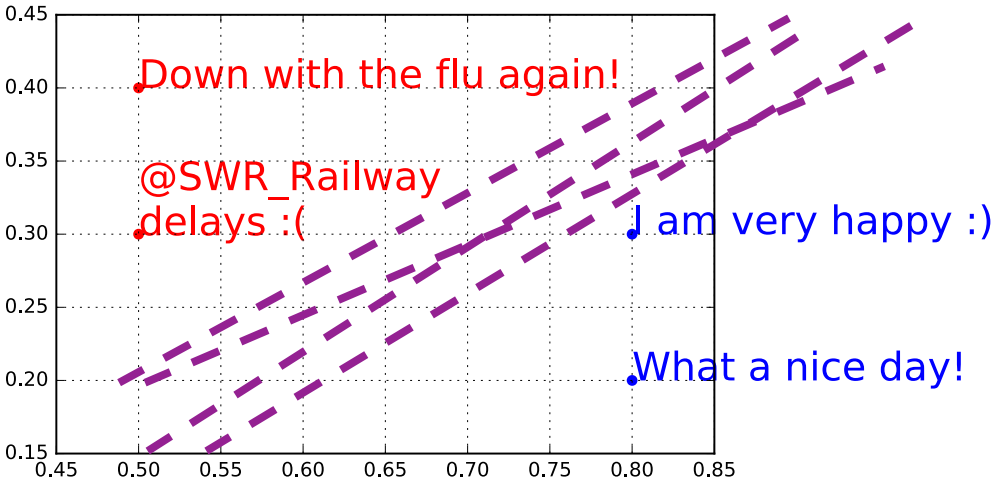
- Linear classifiers try to learn a hyperplane that separates two classes in feature-space with **minimum error**
- Lots of hyperplanes to choose from since different linear classifiers apply different constraints during training

Linear Classifiers



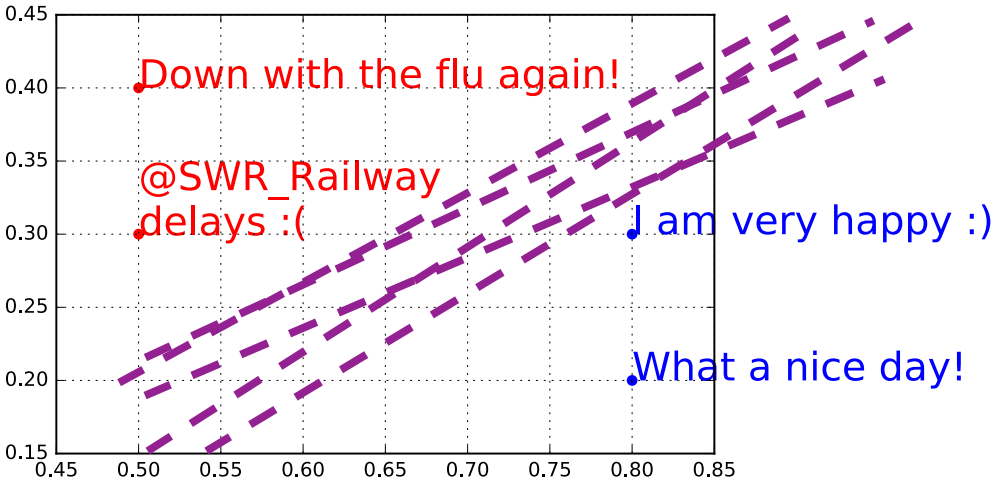
- Linear classifiers try to learn a hyperplane that separates two classes in feature-space with **minimum error**
- Lots of hyperplanes to choose from since different linear classifiers apply different constraints during training

Linear Classifiers



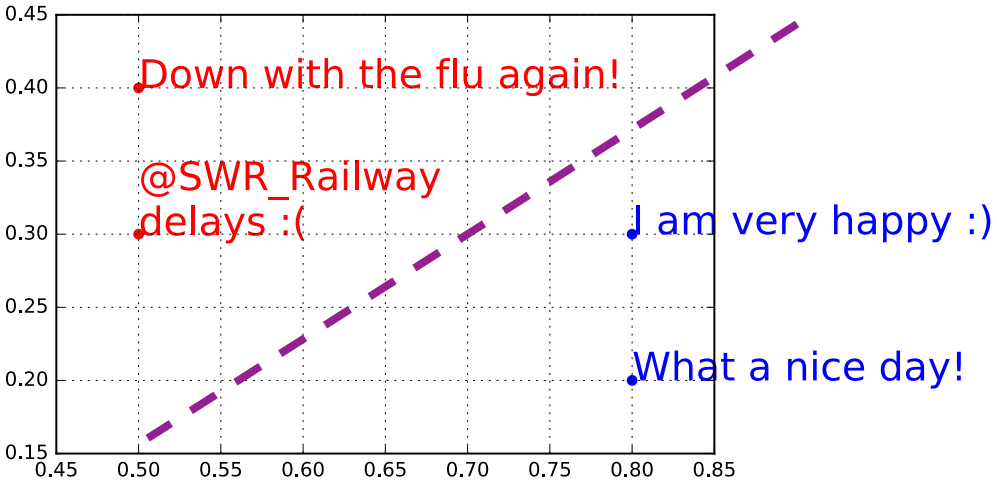
- Linear classifiers try to learn a hyperplane that separates two classes in feature-space with **minimum error**
- Lots of hyperplanes to choose from since different linear classifiers apply different constraints during training

Linear Classifiers



- Linear classifiers try to learn a hyperplane that separates two classes in feature-space with **minimum error**
- Lots of hyperplanes to choose from since different linear classifiers apply different constraints during training

Linear Classifiers



- Linear classifiers try to learn a hyperplane that separates two classes in feature-space with **minimum error**
- Lots of hyperplanes to choose from since different linear classifiers apply different constraints during training

Unsupervised Machine Learning: Clustering

- Clustering aims to **group** data without any prior knowledge of what the groups should look like or contain
- In terms of feature-vectors, items with similar vectors should be grouped together by a clustering operation
- Some clustering operation create **overlapping** groups whilst others are **disjoint** clustering that assign an item to a single group

Unsupervised Machine Learning: Clustering

- Clustering aims to **group** data without any prior knowledge of what the groups should look like or contain
- In terms of feature-vectors, items with similar vectors should be grouped together by a clustering operation
- Some clustering operation create **overlapping** groups whilst others are **disjoint** clustering that assign an item to a single group

Unsupervised Machine Learning: Clustering

- Clustering aims to **group** data without any prior knowledge of what the groups should look like or contain
- In terms of feature-vectors, items with similar vectors should be grouped together by a clustering operation
- Some clustering operation create **overlapping** groups whilst others are **disjoint** clustering that assign an item to a single group

Let's Get Our Hands "Dirty"

https://github.com/pvougiou/WAIS-Away-Day