Supervised Learning

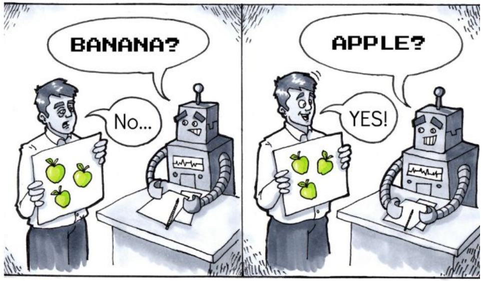
Let's use the first ~20 minutes of the class to finish and review the last class exercises and **materials**.
We will be moving on from those.

We will kick-off today's topics at ~17:50

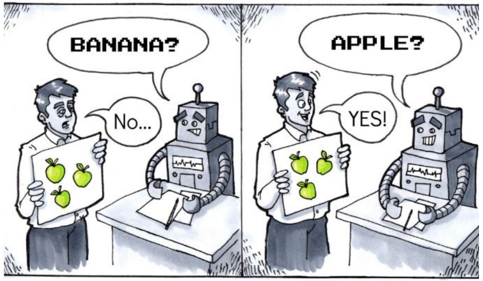
P.s you can move on from store sales and take a look at the wine quality exercises as well if you haven't yet

Machines that Learn

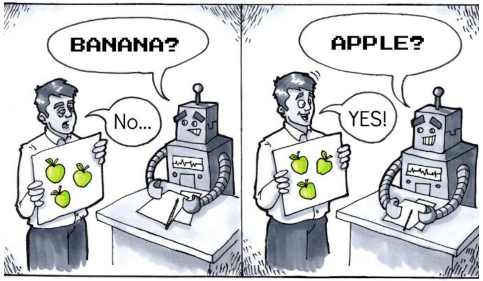
Today.



Supervised Learning



Supervised Learning



Supervised Learning

More classes and missing classes

Supervised Learning

So far.. (mostly)

Regression:

- We've looked at regression problems, where the goal is to predict a continuous output (e.g., predicting house prices).

Classification:

- We've also covered binary classification, where the goal is to predict one of two possible classes (e.g., spam vs. not spam, survived vs not survived, …).

Supervised Learning

So far.. (mostly)

Regression:

- We've looked at regression problems, where the goal is to predict a continuous output (e.g., predicting house prices).

Classification:

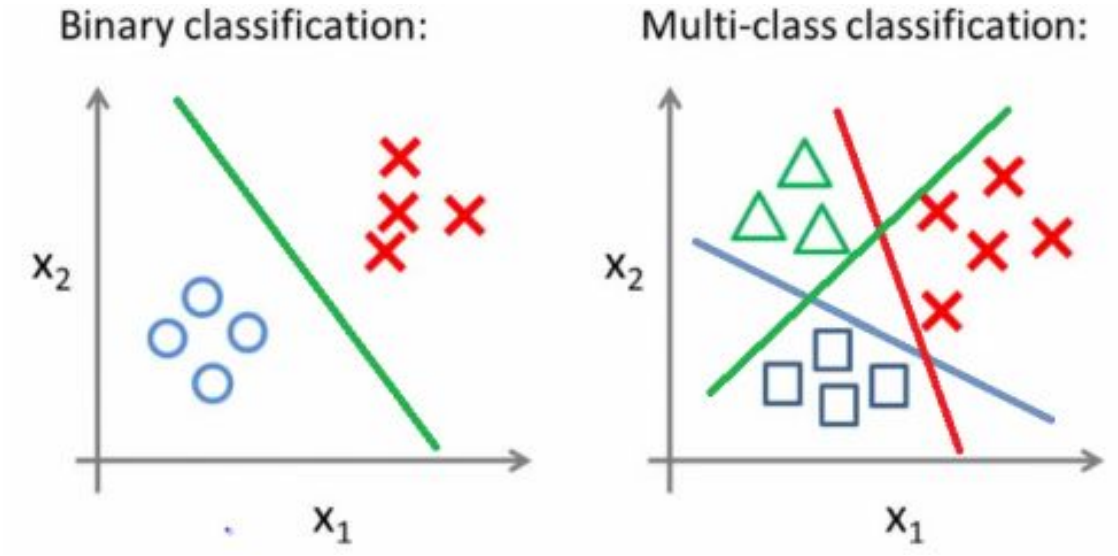
- We've also covered binary classification, where the goal is to predict one of two possible classes (e.g., spam vs. not spam).

Today: Multi-Class Classification:

- Now, we're diving into multi-class classification problems, where the goal is to predict one out of many possible categories (e.g., classifying types of flowers, animal species, various quality of wines, various type of sentiments of a tweet, etc).

Supervised Learning

Multi-Class Classification



A nice read: [sklearn multiclass]

Supervised Learning

Why is it different?

Not all models can *really perform* multi-class out of the box.

Evaluation is more complex.

There’s a higher likelihood of class imbalance.

Supervised Learning

Why is it different?

Not all models can *really perform* multi-class out of the box.

For example Logistic Regression and SVM by themselves only separate two set of points (binary)

Supervised Learning

Why is it different?

Not all models can *really perform* multi-class out of the box.

For example Logistic Regression and SVM by themselves only separate two set of points (binary)

However, KNN, Naive Bayes and Decision Trees* can perform multi-class classification without any extra work.

*more on this later

Supervised Learning

So, we can’t use certain algorithms?

Supervised Learning

So, we can't use certain algorithms?

That's an option.

But any binary algorithm can be transformed into multi-class.

Supervised Learning

So, we can’t use certain algorithms?

That’s an option.

But any binary algorithm can be transformed into multi-class.

Let’s brainstorm.

How could we use a logistic regression to predict 3,4,... classes?

Supervised Learning

So, we can’t use certain algorithms?

Any binary algorithm can be transformed into multi-class.

There are two options:

- One versus One (OvO)
- One versus Rest (OvR)

Supervised Learning

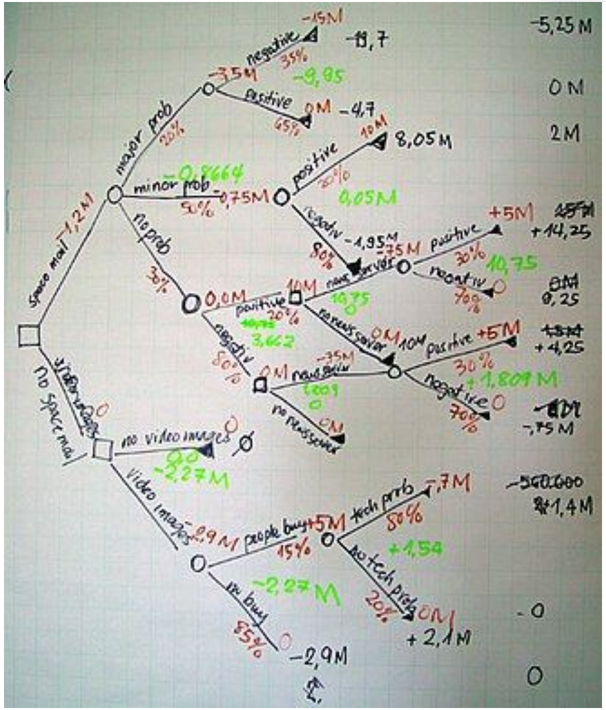
The more classes we have the more important (and also difficult) are the following questions.

What are we optimizing for?
What are we solving?
Which metrics are really useful?
How do we know if we’re doing a good job?
How do we show these results to the business?
What’s the impact of our models?

Supervised Learning

New algorithm alert

Decision Trees



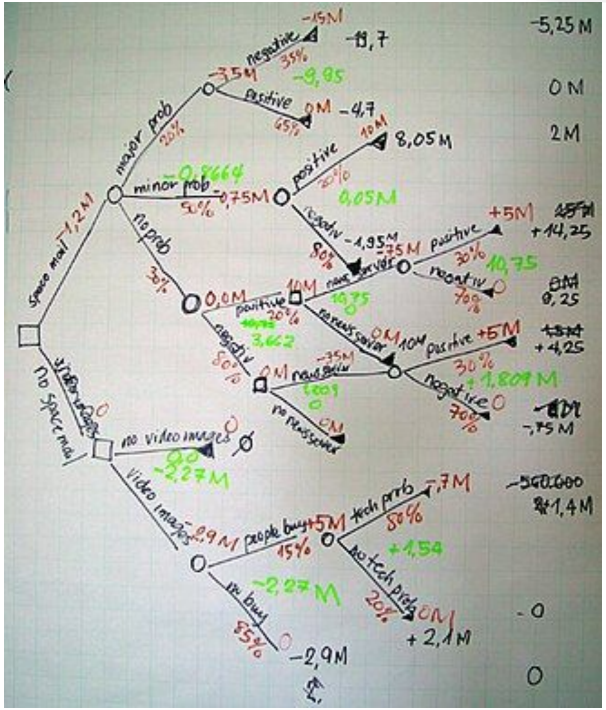
[sklearn decision tree]

Supervised Learning

New algorithm alert

Decision Trees

- Classification
- Regression
- Interpretable



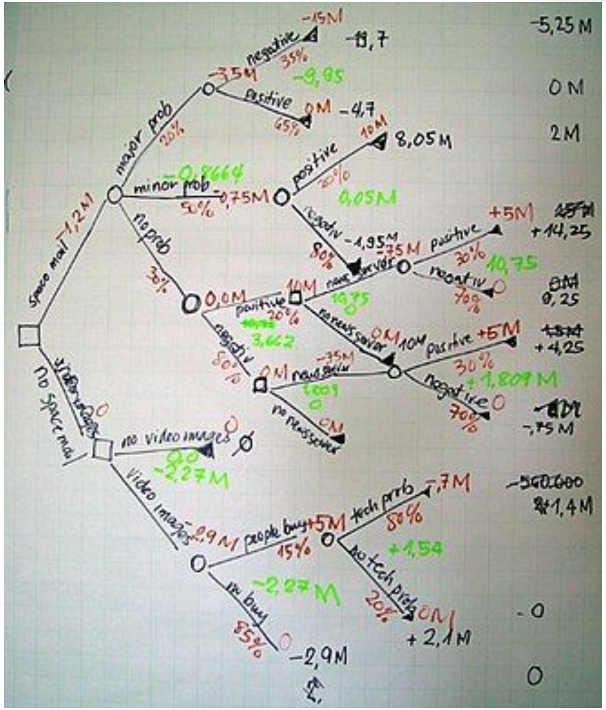
[Decision Trees (IBM Blog Post)]

[sklearn decision tree]

Supervised Learning

Decision Trees

Splits data recursively until it finds a pure leaf node



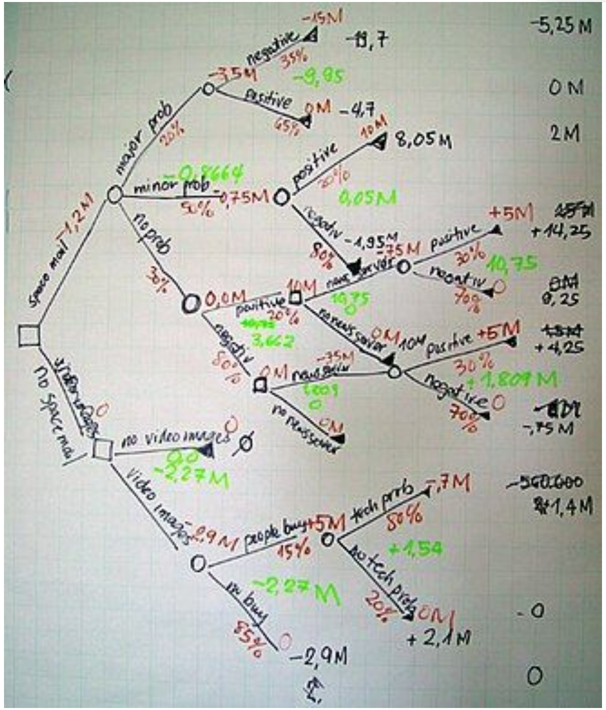
[sklearn decision tree]()

Supervised Learning

Decision Trees

Splits data recursively until it finds a pure leaf node

A decision tree is a **flowchart-like model** that splits data into **branches based on feature tests**, leading to outcomes at leaf nodes. It's intuitive, handles diverse data types, and provides interpretable results, though it may overfit without proper pruning.

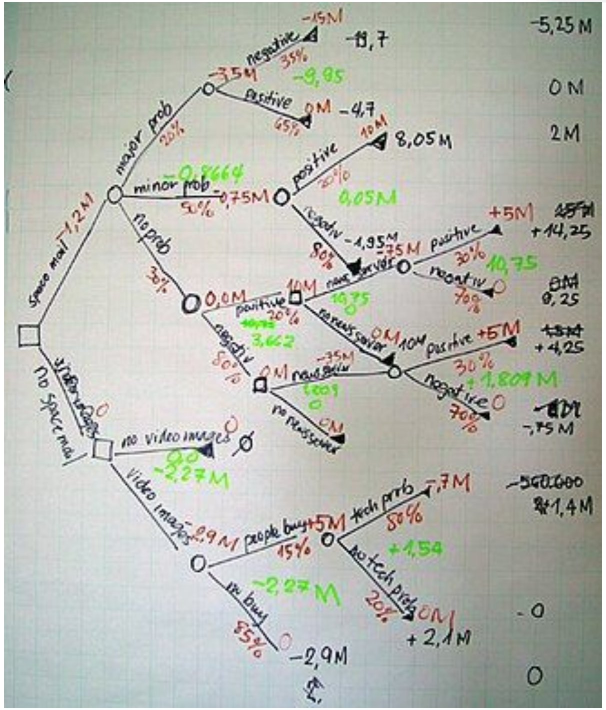


[sklearn decision tree]

Supervised Learning

[Decision Trees]

- Classification
- Regression
- Interpretable



[See example]

[sklearn decision tree]

Supervised Learning

Lets try multi-class.

Use a logistic regression, KNN and Naive-Bayes to predict obesity.

Apply everything you learned so far

- Feature engineering
- Hyperparameter Search and Cross Validation
- Evaluation (Accuracy, Precision and Recall)
  (try using classification_report method)

P.s for the sake of this exercise don't worry too much if you can't get good results at the first try

Supervised Learning

Let’s plant some trees.

Here.

Supervised Learning

Missing data. What do we do?
Not enough classes of X. What do we do?

kaggle example

Supervised Learning

Missing data. What do we do?
Not enough classes of X. What do we do?

- Class Weighting
- Oversampling
- Undersampling

Supervised Learning

Missing data. What do we do?
Not enough classes of X. What do we do?

- Class Weighting

P.s some models are more sensitive than others e.g practical examples

Supervised Learning

Imagine you're training a model to **classify emails** as "**spam**" or "**not spam**," but only **10%** of your emails are spam. Without adjustments, **the model could just predict "not spam" all the time and still be 90% accurate**. However, this wouldn't be helpful because the spam emails (**minority class**) are overlooked.

Class weighting tells the model:

- "Pay more attention to spam emails because there are fewer of them."
- "You don't need to put as much emphasis on non-spam emails since there are so many."

Supervised Learning

class_weight : *dict or 'balanced', default=None*

Weights associated with classes in the form `{class_label: weight}`. If not given, all classes are supposed to have weight one.

The "balanced" mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as `n_samples / (n_classes * np.bincount(y))`.

Note that these weights will be multiplied with sample_weight (passed through the fit method) if sample_weight is specified.

," but

ust

wouldn't

- "You don't need to put as much emphasis on non-spam emails since there are so many."

Supervised Learning

Missing data. What do we do?
Not enough classes of X. What do we do?

- Oversampling
- Undersampling

Supervised Learning

Missing data. What do we do?
Not enough classes of X. What do we do?

- Oversampling (SMOTE)
- Undersampling (Random Under Sampler)

Imbalance learn is your friend

Supervised Learning

Lets try that.

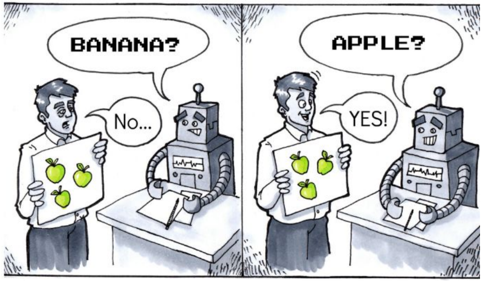
Go back to the obesity prediction task and balance the dataset through class weighting.

Does it improve the results?

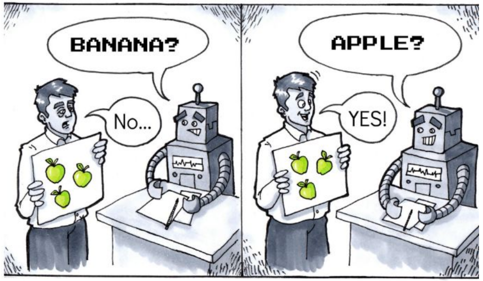
What if you undersample the data?

Machines that Learn

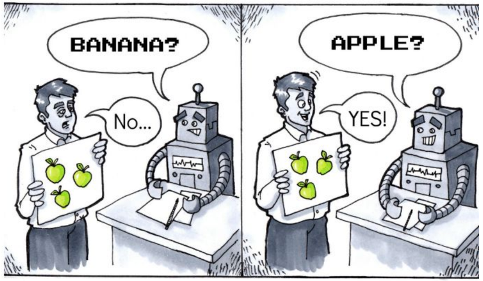
More on the next episode.



Supervised Learning



Supervised Learning



Supervised Learning

Evaluation Metrics (again, yes it's important)
Ensembles.