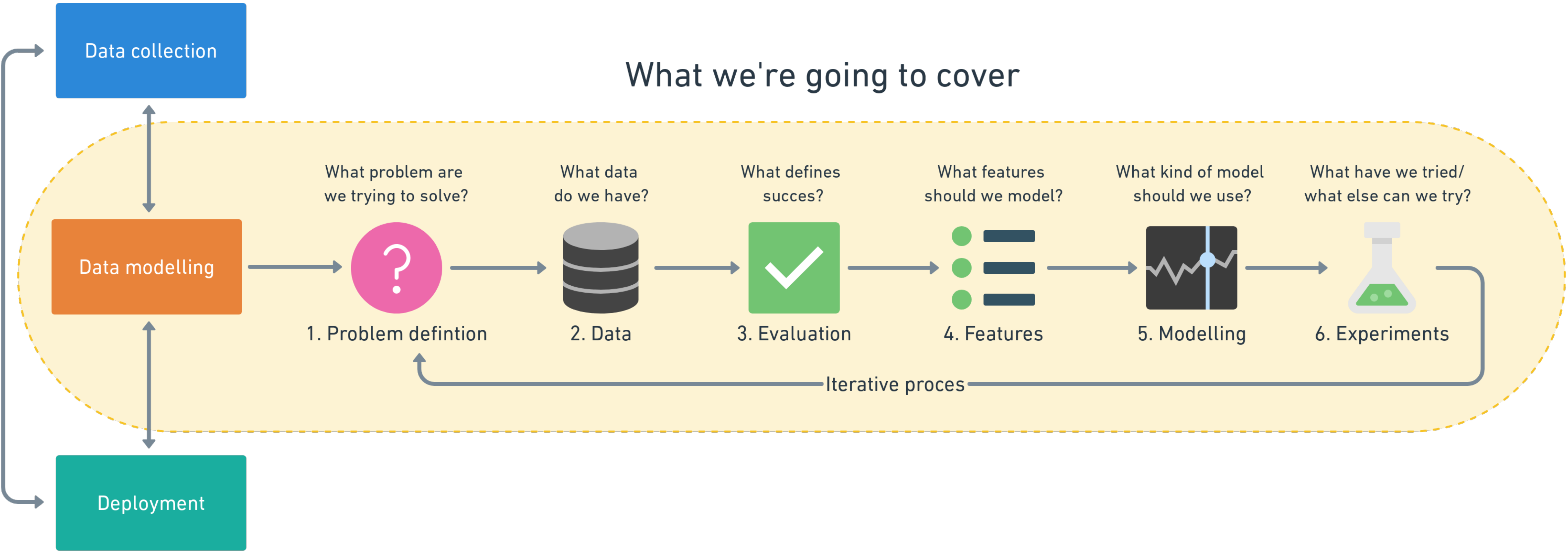
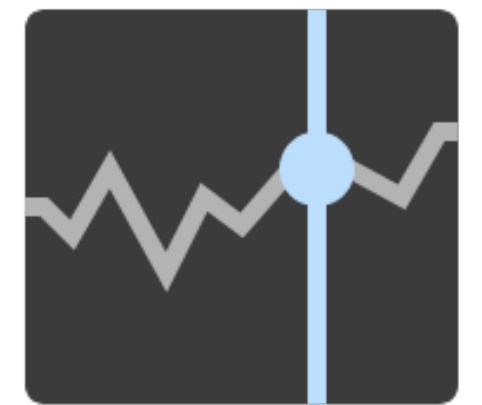


Steps in a full machine learning project



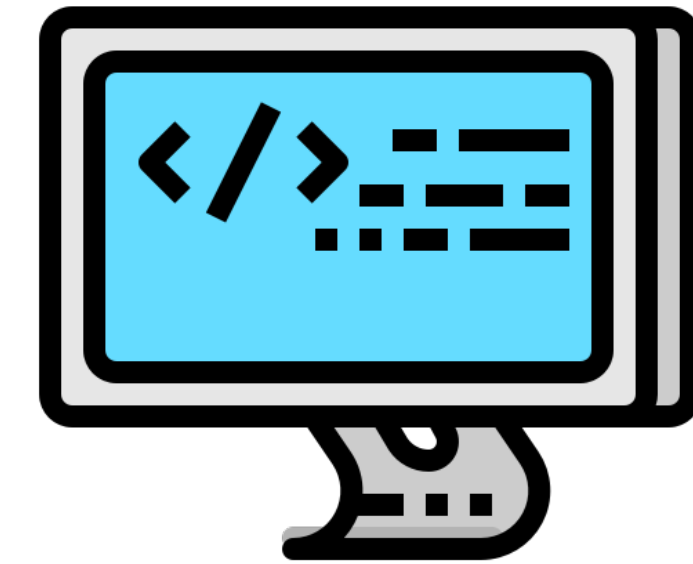
5. Modelling Part 1 — 3 sets



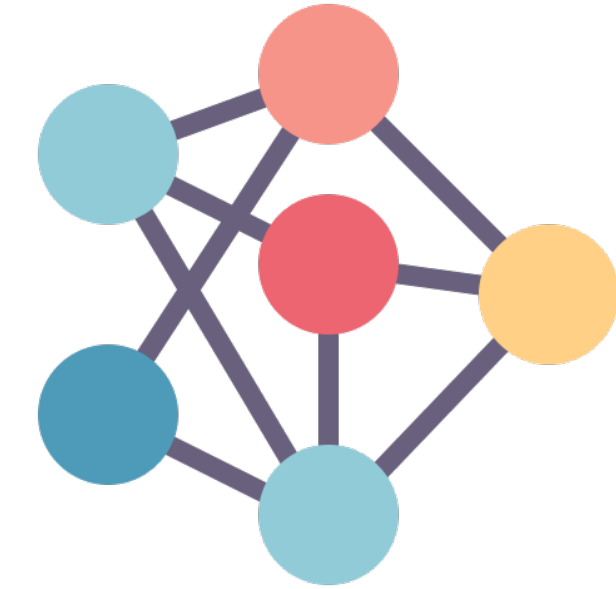
“Based on our problem and data, what model should we use?”

3 parts to modelling

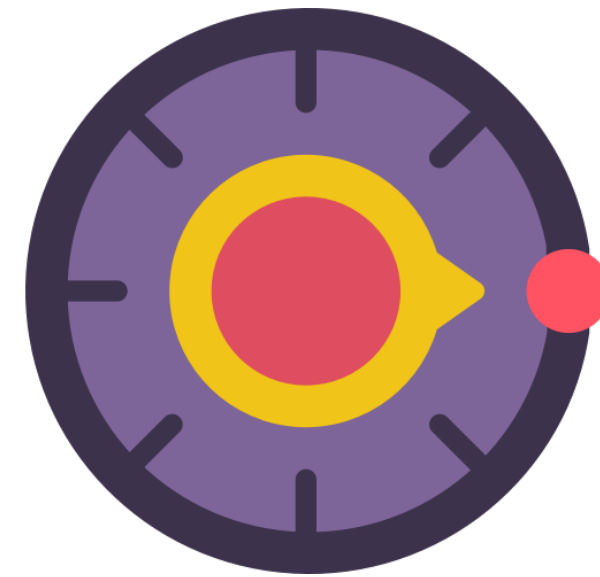
1. Choosing and training a model



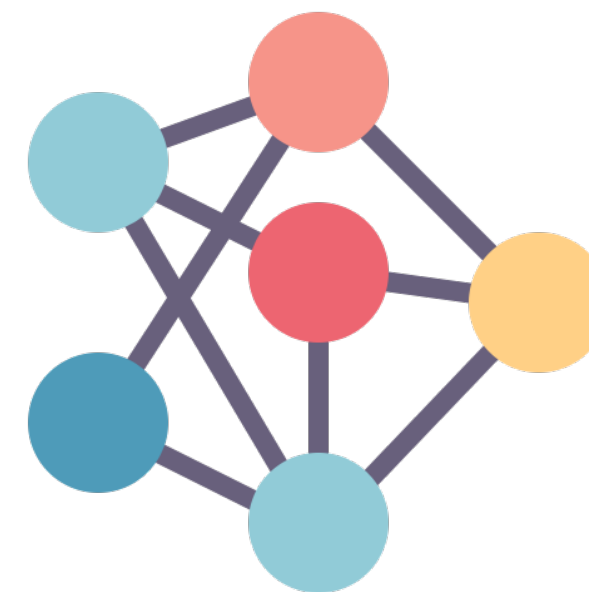
or



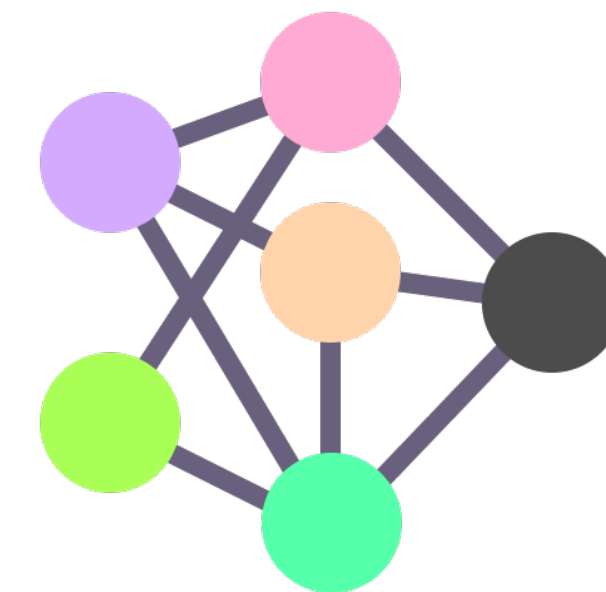
2. Tuning a model



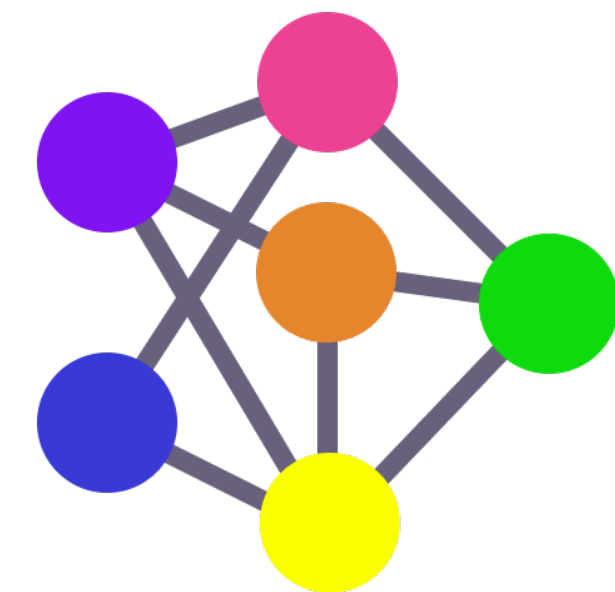
3. Model comparison



vs.



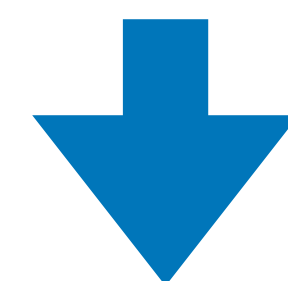
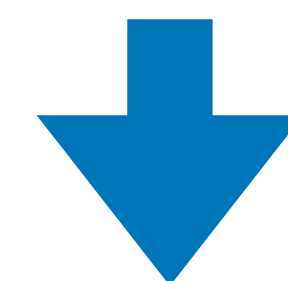
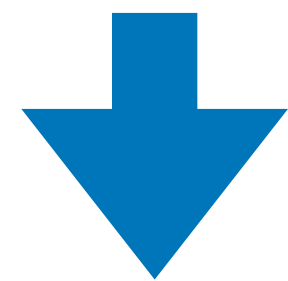
vs.



The most important concept in machine learning

(the training, validation and test sets or 3 sets)

Your data



Split



Training

Validation

Test

Train your model on this

Tune your model
on this

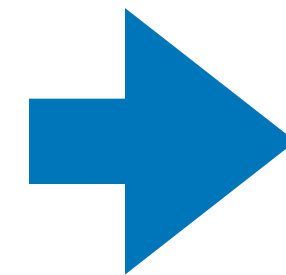
Test and
compare on this

The most important concept in machine learning

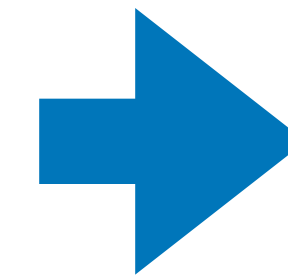
(the 3 sets)



Course materials
(training set)



Practice exam
(validation set)

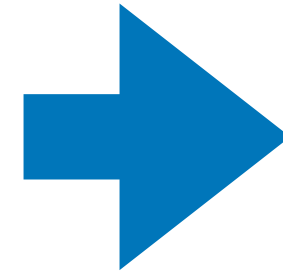


Final exam
(test set)

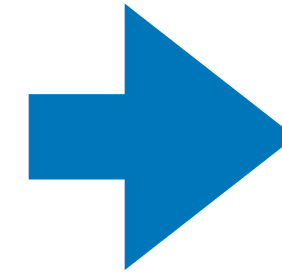
Generalization

The ability for a machine learning model to perform well on data it hasn't seen before.

When things go wrong



Practice exam
(same as final exam set)

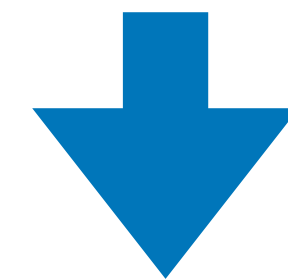
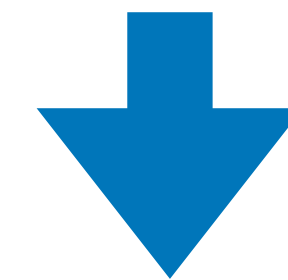
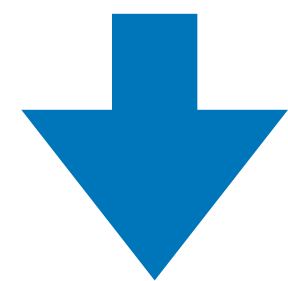


Final exam
(already seen it)



The most important concept in machine learning

(the 3 sets)



Split



Training split (70-80%)

Validation split
(10-15%)

Test split
(10-15%)

**What was the last thing you testing
your ability on?**