Validation and tuning in ML applications

Ron Wehrens

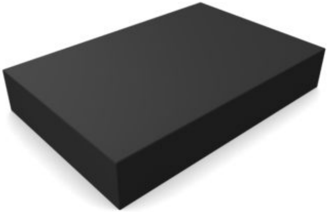
October 26, 2020

Biometris
Wageningen University & Research

Goals

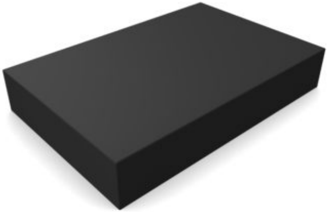
- provide general principles
- perhaps take away some misconceptions
- hear what you think

Training a ML model



- take a model off the shelf
- train it with data
- make predictions

Training a ML model



- take a model off the shelf
- train it with data
- make predictions

How good is it?

- predict unseen data

Evaluating your model when data are scarce

Model training: use all available data

Evaluating your model when data are scarce

Model training: use all available data

Validation:

- divide your data in k segments
- fit k models, leaving out each segment once
- predict the left out segments
- aggregate prediction errors
- Crossvalidation (CV)

Evaluating your model when data are scarce

Model training: use all available data

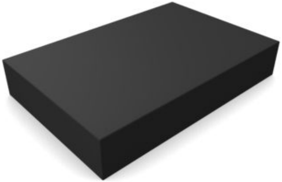
Validation:

- divide your data in k segments
- fit k models, leaving out each segment once
- predict the left out segments
- aggregate prediction errors
- Crossvalidation (CV)

CV result: an error estimate

But!

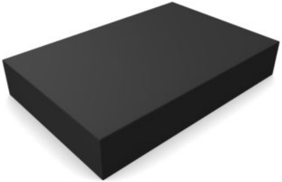
As is, off the shelf



Only few methods:
LDA, QDA, LS regression,
but also pretrained networks
like Mask R-CNN, YOLO, ...

But!

As is, off the shelf



Only few methods:
LDA, QDA, LS regression,
but also pretrained networks
like Mask R-CNN, YOLO, ...

Tuning required!

Virtually all ML methods are like this - unfortunately!

Tuning ML models

Solution: pretty simple

- Define the sets of tuning parameters
- For each, train a model
- Evaluate all models using CV
- Pick the best one(s)

Tuning ML models

Solution: pretty simple

- Define the sets of tuning parameters
- For each, train a model
- Evaluate all models using CV
- Pick the best one(s)

Be sensible: small differences are unlikely to be significant

Tuning ML models

Solution: pretty simple

- Define the sets of tuning parameters
- For each, train a model
- Evaluate all models using CV
- Pick the best one(s)

Be sensible: small differences are unlikely to be significant

Other model selection criteria are possible, too: AIC, BIC, Cp, ...

But!

What about the prediction error?

- CV estimates are optimistically biased!

But!

What about the prediction error?

- CV estimates are optimistically biased!
- You need additional data to evaluate the true performance

But!

What about the prediction error?

- CV estimates are optimistically biased!
- You need additional data to evaluate the true performance
- or... double crossvalidation

Training and validation strategy

- Set apart a test data set

Training and validation strategy

- Set apart a test data set
- The rest will be used for training

Training and validation strategy

- Set apart a test data set
- The rest will be used for training
- Set up your models using all training data

Training and validation strategy

- Set apart a test data set
- The rest will be used for training
- Set up your models using all training data
- Validate your models using CV (training data only!!!)

Training and validation strategy

- Set apart a test data set
- The rest will be used for training
- Set up your models using all training data
- Validate your models using CV (training data only!!!)
- Pick your favourite model

Training and validation strategy

- Set apart a test data set
- The rest will be used for training
- Set up your models using all training data
- Validate your models using CV (training data only!!!)
- Pick your favourite model
- Estimate the prediction error of your favourite model by predicting the unseen test data

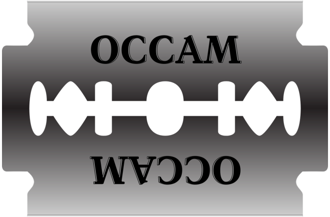
Simple. What can go wrong?

Lots of things.

Simple. What can go wrong?

Lots of things.

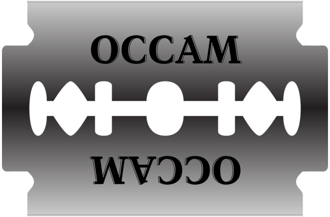
- Scaling
- Unfortunate choice of test data
- Using hundreds of different models
- Using unnecessary complicated models
- ... and many other things



Simple. What can go wrong?

Lots of things.

- Scaling
- Unfortunate choice of test data
- Using hundreds of different models
- Using unnecessary complicated models
- ... and many other things



Moral of the story: keep it simple, stupid!

Take home messages

- Avoid. More. Than. One. Tuning. Parameter.
  - as found in, e.g., SVMs, neural nets, the elastic net, ...

Take home messages

- Avoid. More. Than. One. Tuning. Parameter.
  - as found in, e.g., SVMs, neural nets, the elastic net, ...
- Know your data!
  - exploratory analysis
  - mind the scaling

Take home messages

- Avoid. More. Than. One. Tuning. Parameter.
  - as found in, e.g., SVMs, neural nets, the elastic net, ...
- Know your data!
  - exploratory analysis
  - mind the scaling
- Know your question!
  - enough is enough

Take home messages

- Avoid. More. Than. One. Tuning. Parameter.
  - as found in, e.g., SVMs, neural nets, the elastic net, ...
- Know your data!
  - exploratory analysis
  - mind the scaling
- Know your question!
  - enough is enough
- Easy to learn, hard to master