COMS30035, Machine learning: Combining Models 1, Selection

Edwin Simpson
edwin.simpson@bristol.ac.uk

Department of Computer Science, SCEEM
University of Bristol

November 14, 2023

Agenda

- Model Selection
- Model Averaging
- Ensembles: Bagging
- Ensembles: Boosting and Stacking
- Tree-based Models
- Conditional Mixture Models
- Ensembles of Humans

Textbook

We will follow Chapter 14 of the Bishop book: Bishop, C. M., Pattern recognition and machine learning (2006). Available for free here.

The Model Selection Problem

- Select a model h from a set of possible models in a set H,

The Model Selection Problem

- Select a model h from a set of possible models in a set H,
- The models in H can differ in various ways, such as:

The Model Selection Problem

- Select a model h from a set of possible models in a set H,
- The models in H can differ in various ways, such as:
  - The structure of the model, e.g., differences in graphical models;

The Model Selection Problem

- Select a model h from a set of possible models in a set H,
- The models in H can differ in various ways, such as:
  - The structure of the model, e.g., differences in graphical models;
  - The choice of prior distribution $p(\theta)$ over model parameters;

The Model Selection Problem

- Select a model h from a set of possible models in a set H,
- The models in H can differ in various ways, such as:
  - The structure of the model, e.g., differences in graphical models;
  - The choice of prior distribution $p(\theta)$ over model parameters;
  - The learning algorithm used, e.g., EM, MCMC, backpropagation,;
  - Parameters of the learning algorithm, like learning rate for stochastic gradient descent (SGD);

The Model Selection Problem

- Select a model h from a set of possible models in a set H,
- The models in H can differ in various ways, such as:
  - The structure of the model, e.g., differences in graphical models;
  - The choice of prior distribution $p(\theta)$ over model parameters;
  - The learning algorithm used, e.g., EM, MCMC, backpropagation,;
  - Parameters of the learning algorithm, like learning rate for stochastic gradient descent (SGD);
  - The features of the data used as inputs to the model;

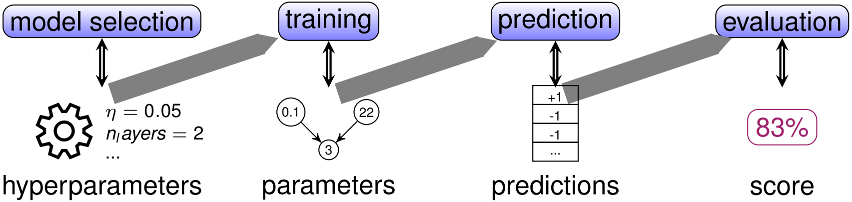
The Model Selection Problem

- Select a model h from a set of possible models in a set H,
- The models in H can differ in various ways, such as:
  - The structure of the model, e.g., differences in graphical models;
  - The choice of prior distribution $p(\theta)$ over model parameters;
  - The learning algorithm used, e.g., EM, MCMC, backpropagation,;
  - Parameters of the learning algorithm, like learning rate for stochastic gradient descent (SGD);
  - The features of the data used as inputs to the model;
  - The examples in the dataset the model is trained on;

The Model Selection Problem

- Select a model h from a set of possible models in a set H,
- The models in H can differ in various ways, such as:
  - The structure of the model, e.g., differences in graphical models;
  - The choice of prior distribution $p(\theta)$ over model parameters;
  - The learning algorithm used, e.g., EM, MCMC, backpropagation,;
  - Parameters of the learning algorithm, like learning rate for stochastic gradient descent (SGD);
  - The features of the data used as inputs to the model;
  - The examples in the dataset the model is trained on;
  - Random initialisation of parameters for EM or SGD

Hyperparameters



- It's useful to characterise all of these modelling decisions as *hyperparameters*
- Hyperparameters = all modelling choices that are fixed before training

Model Selection on a Validation Set

Dataset:

- Train a set of models H with different hyperparameters
- Choose the model h that maximises performance on a validation set.
  - Can't tune on the training set as it would lead to overfitting

Model Selection on a Validation Set

Dataset:

- Train a set of models H with different hyperparameters
- Choose the model h that maximises performance on a validation set.
  - Can't tune on the training set as it would lead to overfitting
- Strength: optimises a performance metric we really care about.

Model Selection on a Validation Set

Dataset: | Train | Validation /Development | Test

- Train a set of models H with different hyperparameters
- Choose the model h that maximises performance on a validation set.
  - Can't tune on the training set as it would lead to overfitting
- Strength: optimises a performance metric we really care about.
- Weakness: we didn't use all the data in training;

Edwin Simpson
edwin.simpson@bristol.ac.uk

Model Selection on a Validation Set

Dataset:

- Train a set of models *H* with different hyperparameters
- Choose the model *h* that maximises performance on a validation set.
  - Can't tune on the training set as it would lead to overfitting
- Strength: optimises a performance metric we really care about.
- Weakness: we didn't use all the data in training;
- Weakness: if the validation set is small, we might choose the wrong model!

Edwin Simpson
edwin.simpson@bristol.ac.uk

Model Selection on a Validation Set

Dataset:	Train	Validation /Development	Test

- Train a set of models H with different hyperparameters
- Choose the model h that maximises performance on a validation set.

Model Selection on a Validation Set

Dataset:	Train	Validation /Development	Test

- Train a set of models H with different hyperparameters
- Choose the model h that maximises performance on a validation set.
- How do we come up with H?

Model Selection on a Validation Set

Dataset:

- Train a set of models H with different hyperparameters
- Choose the model h that maximises performance on a validation set.
- How do we come up with H?
- Random search: test random combinations of hyperparameters

Edwin Simpson
edwin.simpson@bristol.ac.uk

Model Selection on a Validation Set

Dataset: Train | Validation /Development | Test

- Train a set of models *H* with different hyperparameters
- Choose the model *h* that maximises performance on a validation set.
- How do we come up with *H*?
- Random search: test random combinations of hyperparameters
- Grid search: For each hyperparameter, define a set of values to test
  - Use your knowledge of the problem to test only reasonable values
  - For numerical hyperparameters, e.g., learning rate, choose a set of evenly-spaced values within a sensible range
  - *H* contains all combinations of the chosen hyperparameter values

Edwin Simpson
edwin.simpson@bristol.ac.uk

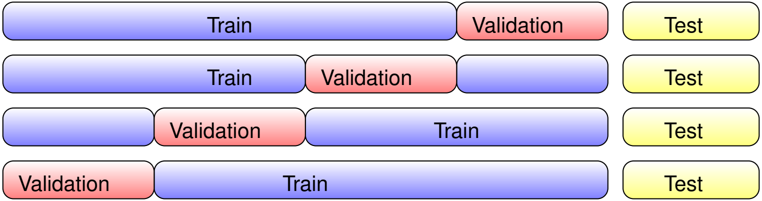
Model Selection on a Validation Set

Dataset: Train | Validation /Development | Test

- Train a set of models H with different hyperparameters
- Choose the model h that maximises performance on a validation set.
- How do we come up with H?
- Random search: test random combinations of hyperparameters
- Grid search: For each hyperparameter, define a set of values to test
  - Use your knowledge of the problem to test only reasonable values
  - For numerical hyperparameters, e.g., learning rate, choose a set of evenly-spaced values within a sensible range
  - H contains all combinations of the chosen hyperparameter values
- Reduce the number of tests needed to find a good combination using a more intelligent strategy such as Bayesian Optimisation

Edwin Simpson
edwin.simpson@bristol.ac.uk

Cross-validation



- Split the training data into k random, equally-sized subsets;
- For each of the k folds: leave out the kth subset from training, train on the rest and test on the kth subset;
- Compute the average performance across all k folds;
- Avoids overfitting by tuning on training set performance...
- And avoids tuning on a single small validation set.

Now do the quiz!

Please do the quiz for this lecture on Blackboard.