4F13: Probabilistic Machine Learning

Summarized from C. Rasmussen & D. Krueger lectures, Michaelmas 2021

Oussama Chaib

October 2021

Contents

1 Modelling data

1.1 Purpose of models

The purpose of models is:

- Making predictions
- Generalizing: interpolation, extrapolation
- Generating more data from a similar distribution as the training set
- Compressing and summarizing data
- Interpreting statistical relationships in data
- Evaluating the relative probability of a hypothesis on data

1.2 Origin of models

The origin of models can be:

- **First principles:** (i.e: Newtonian mechanics model, high level of accuracy)
- **Observations and data:** (i.e: annual production of timber depending on climate and geographical factors)

Definition – Machine learning is a broad term that covers theory and practice of mathematical models which to a significant degree rely on data.

1.3 Priors

Every model relies on priors:

- **Knowledge**
- **Assumptions** (could be true or false)
- **Simplifying assumptions** (not necessarily true, but good enough – i.e: the mistake associated with the assumption is fairly small even though it might not be necessarily true)

1.4 Components of a model

Time series have:

- Unobserved/hidden/latent variables ($x(t)$, $x(t-1)$)
- Observations (shaded $y(t)$, $y(t-1)$)
- Parameters to link everything
  - Transitions (between latent variables)
  - Emissions (from a lantent variable to an observation)

<u>*Note:*</u> The number of latent variables increases with the number of observations, but the number of parameters doesn't!
– **Learning/training models:** "What to do with all this data?"
Depending on the data, some models include: inference, estimation, sampling, and marginalization.

1.5 Practical modelling

1. Treat (training) the unobserved quantities (latent variables, observations, parameters)
2. Make predictions based on test cases, interpret the trained model (can we figure out what the model is trying to tell us about the data?)
3. Evaluate the accuracy of the data
4. Model selection and criticism (choose the right model or variant of the model, identify limitations)

There is not "true" or "correct" model – *"All models are wrong, but some are useful"* - George E.T. Box

2 Linear in the parameters regression

Let's start off with a dataset $D = x_i, y_i{}_{i=1}^{N}$. From a dataset of N points, we would like to infer the coordinate y_* at x_*. A simple model to do that would be polynomial **linear in the parameters** regression:

$$f_w(x) = w_0 + w_1 x + w_2 x^2 + ... + w_M x^M$$

where w_i are the corresponding weights of the polynomial, and the **parameters** of the model.
– Relevant questions:
Model structure: Should we choose a polynomial? What degree M should we choose? *Parameters:* What values of w_i do we choose?

2.1 Least squares approach

Let's find the "best" polynomial (degree M and weights w_i) according to the least squares approach (minimizing the variance or sum of squared error e_i^2):

$$e_i(x)^2 = (y_i(x_i) - f_w(x_i))^2$$

$$E(x) = \sum_{i=1}^{N} e_i^2$$

3 Likelihood and noise

3.1 Comments from QnA

– Why is the error vector e minimal if it's orthogonal to all columns of ϕ?
ϕ is a fixed function, it doesn't have any parameters in it. Still linear in parameters because the product of w and ϕ matrices is linear (even at high polynomial orders).
– What is Euclidian geometry?
Simple geometries, straight lines, basic shapes (including circle).
– Bayesian methods, why are they not as popular?:
Maximum likelihood (used often, single value that best explains the data, quite popular/successful).