Lecture I – What are Non-Parametric models?

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

STAT/BIOST 527
Spring 2023

Why non-parametric?

Why non-parametric?

- neural networks (nn)
- Random Forests (RF)
- Gaussian Processes (GP)
- Support Vector Machines/Kernel machines (SVM)

  . . . are all non-parametric

Why non-parametric?

- neural networks (nn)
- Random Forests (RF)
- Gaussian Processes (GP)
- Support Vector Machines/Kernel machines (SVM)

  . . . are all non-parametric

simple, convenient models are also non-parametric

- k-Nearest Neighbors (k-NN)
- kernel regression and kernel density estimation (KDE)

other areas of non-parametric statistics

- Non-linear dimension reduction/Manifold learning (NLDR)

Non-parametric models...

Are useful when

- lots of data available (n large)
- lots of computing power available

NP models/NP statistics

- need more data, more computer memory, are slower to use, are harder to analyze than parametric models
- BUT they **“adapt to the shape of the data”**
- the NP model’s **level of detail** increases with n

- allow us to model a data distribution with arbitrary accuracy (if n is large enough, and we know what we are doing)

Parametric vs. non-parametric models

Example 1 (Parametric and non-parametric predictors)

Parametric

- Linear, logistic regression
- Linear Discriminant Analysis (LDA)
- Neural networks (few parameters)
- Naive Bayes
- CART with L levels

Non-parametric

- Nearest-neighbor classifiers and regressors
- Nataraya-Watson kernel predictors
- Monotonic regression
- Neural networks (large, deep)
- Random Forests
- Support Vector Machines
- Gaussian process regression
-

A mathematical definition

- A model class $\mathcal{F}$ is **parametric** if it is finite-dimensional, otherwise it is **non-parametric**

In other words

- When we estimate a parametric model from data, there is a fixed number of parameters, (you can think of them as one for each dimension, although this is not always true), that we need to estimate to obtain an estimate $\hat{f} \in \mathcal{F}$.
- The parameters are meaningful.
  E.g. the β_j in logistic regression has a precise meaning: the component of the normal to the decision boundary along coordinate j.
- The dimension of β does not change if the sample size n increases.

Non-parametric models – Some intuition

- When the model is non-parametric, the model class $\mathcal{F}$ is a function space.
- The $\hat{f}$ that we estimate will depend on some numerical values (and we could call them parameters), but these values have little meaning taken individually.
- The number of values needed to describe $\hat{f}$ generally grows with n. Example In the Nearest neighbor and kernel predictors, we have to store all the data points, thus the number of values describing the predictor f grows (linearly) with the sample size.
- Non-parametric models often have a **smoothness parameter**.
  Example Smoothness parameters K in K-nearest neighbor, h the kernel bandwidth in kernel regression.
  To make matters worse, a smoothness parameter is not a parameter! More precisely it is not a parameter of an $f \in \mathcal{F}$, because it is not estimated from the data, but a descriptor of the model class $\mathcal{F}$.

Do NP models have parameters?

Parametric model

- p parameters $= \dim \mathcal{F}$
- parameters "have meaning"
- parameters are estimated from data

Non-parametric model

- **smoothness parameters**
  - not estimated from data
  - similar to hyperparameters in Bayesian statistics
  - can be selected by Cross-validation, etc.

regular parameters

- usually estimated from data
- have little interpretation on their own
- their number (sometimes) depends on the sample size and data distribution

Examples of regular parameters

- all data set for k-NN
- support vectors for SVN
- weights for a nn
- splits of trees for a random forest (RF)

Examples of smoothing parameters

- k for k-NN
- kernel width (or kernel parameters) for SVN/kernel machines
- number of leaves for a random forest (RF)

This course

I Basics

- Nearest neighbors regression, classification, density estimation (k-NN)
- Kernel regression, classification, density estimation
- Non-parametric clustering – density based
- (approximate nearest neighbors in big data, high dimensions)
  (cross-validation (CV))

II Getting serious

- Double descent
- Kernel machines
  (fun with kernels for strings, trees, and so on)
- Random Forests
- Manifold learning

III Advanced

- Gaussian processes (GP)
- NP-Bayes (non-parametric clustering)
- nn as GP and the Neural-Tangent Kernel (NTK)
  (shape constrained estimation)
  (NP boostrap)
  (Conformal Prediction)