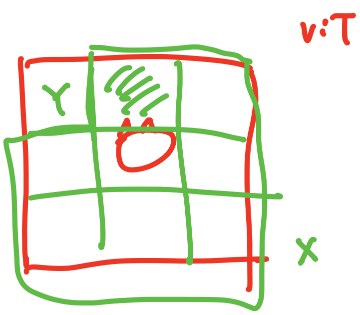
Advanced topics on large foundation models

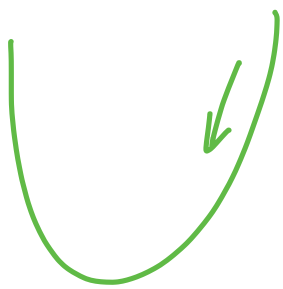
estimation / prediction / uncertainty

inferential analysis
- asymptotic theory.
- bootstrapping (simulation)

predictive analysis (supervised)
1. regression $\underset{\sim}{Y} \leftrightarrow X$
   continuous
2. classification
   Y is discrete
3. generalized linear model

exploratory analysis (unsupervised)
1. clustering (finite mixture models)
2. dimension reduction (latent variable model)





general principles

1) likelihood principle

parametric statistical models
(agnostic setting)

2) Concentration principle $\underbrace{X_1 \cdots X_n} \sim p$

$$\min_{\theta \in \Theta} \frac{1}{n}\sum_{i=1}^{n} \ell(x_i;\theta) \qquad \min_{\theta \in \Theta} \mathbb{E}\,\ell(x;\theta) \qquad x \sim p$$

empirical population

$n \to \infty$
diff $\to 0$

3) regularization principle

restrictions on θ (constraints)
augmentations on ℓ (penalty)
... on x_i

Basics

① statistical modeling $\{P_\theta,\ \theta \in \Theta\}$

② parametric estimation $\hat{\theta}_n(x_1 \cdots x_n) \xrightarrow{\text{as } n\to\infty} \theta$

③ prediction. $(x_i, y_i) \to \hat{\theta}_n(\{x_i, y_i\}_{i=1}^{n})$

④ uncertainty $\to$ a new x
predict the corresponding Y

Def: a statistical model

A set of prob distribution indexed by a parameter θ

$$\mathcal{P} \triangleq \{ P_\theta,\ \theta \in \Theta \}$$

parameter parameter space

① parametric model Θ is finite-dimensional
example: one dimensional Gaussian

② nonparametric model Θ is infinite-dimensional example: sobolev space

$$\{p: \int p(t)\,dt = 1,\ p(t) \geq 0,\ \int p''(t)^2\,dt < \infty\}$$

---

parameter estimation

1. maximum likelihood estimation
2. regualization
3. empirical risk minimization

MLE

Def: $X_1 \cdots X_n \sim p_\theta(x)$

likelihood related to a single datapoint x_i is $\underbrace{L(x_i;\theta)} \triangleq p_\theta(x_i)$.

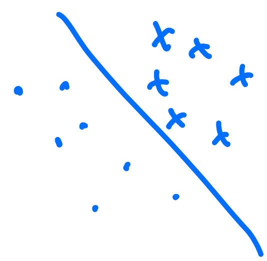
Def: likelihood related to the entire dataset $\{x_i\}_{i=1}^n$

$$L_n(\theta) \triangleq p_\theta(x_1, \dots, x_n)$$

Def: MLE $\hat{\theta}_n = \underset{\theta \in \Theta}{\arg\max}\ L_n(\theta)$

remark: MLE does not exist

(e.g.. logistic regression when data are perfectly separable)



Def: log-likelihood $\ell_n(\theta) \triangleq \underline{\log L_n(\theta)}$