why study stat. learning?

↳ Business: use data collected from daily ops. to ↑ efficiency, predict val., or explore relationships

Δ Developing accurate model to pred. sales from 3 media budgets → pg. 15

- When trying to model 1 var. w/ p other vars., we assume a relationship btwn. target (Y) & features ($X = (X_1, \ldots, X_p)$)

↳ $Y = f(X) + \epsilon$ → Random error → ind. from X & $\mu = 0$

→ Arbitrary function → what info X provides abt. Y

What's the difference between prediction & inference?

PREDICTION	INFERENCE
- Set of X but not their Ys	
- Don't need highly interpretable models	
- Predict Y given X: $\hat{Y} = \hat{f}(X)$	
Δ Predict a patient's risk of reaction from blood characteristics	

PREDICTION

- Reducible error is error caused by est.

$f(x)$

- Irreducible error is simply caused by existence of rand. error
- Improve prediction by reducing reducible error:
  - How? → select proper model
    - tune model

INFERENCE

How do we est. f ?

- ALWAYS assume that we have obs. set of n data pts.
  - Data used to Train / teach method how to estimate f
    - = training data
  - w/ multiple (p) features & n obs. we have data:

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \text{ where}$$

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$$

Parametric vs Non-Parametric Methods ?

- PARAMETRIC: involves 2-step model-based approach:
  - ① Make assumption about SHAPE of f (relationship btwn. y & X) ⇒ model will have a functional form that you can write
    - Δ Linear Regression

↳ ② After selecting a model, use some criteria that uses the training data to fit/train the model

↳ Δ Est. $\beta_0, \beta_1, \ldots, \beta_p$ using OLS

PROS	CONS
- Simplifies model fitting process	- Chosen form is almost never going to be exactly right → risk having poor estimates

How to adress cons of parametric models?

- Choose flexible models that fit many functional forms

↳ BUT makes fitting model more complex → ↑ parameters & ↑ complex models risk overfitting

↳ Models works well on train but doesn't generalize well → follow ϵ too closely

Parametric example: $\text{Income} = \beta_0 + \beta_1 \text{Education} + \beta_2 \text{Seniority}$

↳ Doesn't do a perf. job but may be good enough/reasonable

↳ May be best one can do w/ small data

NON-PARAMETRIC : NO explicit assumptions about form of f

↳ Instead use "algo" or serious of procedures to get as close to data pts. w/o being too rough or wiggly

PROS

- No risk of getting functional form wrong

↳ Can be very accurate

CONS

- Unlike parametric models, there is no simplification → ↑↑ complexity ⇒ requires a lot of obs. to be accurate

Non-Parametric Example: Thin plate spline for Income → tries to est. $\hat{y}$ by getting as close as possible to data while keeping hyperplane smooth

↳ Analyst has to decide on smoothness → if too high the model could fit perf. ⇒ leads to overfitting

PREDICTION ACCURACY VS MODEL INTERPRETABILITY

- Models can be plotted on a range of flexibility
  - ↳ ↑ flexible: Can fit many shapes of data to est. f
    - Δ Splines
  - ↳ ↓ flexible: More restricted in shapes that it can est. for f
    - Δ Linear Regression

Why ever choose a less flexible model?

① If interested in inference → restrictive models = ↑ interpretable

↳ LR allows for easy interpretation of coeff. vs something like splines

② Even when only concerned w/ prediction, ↓ flexible models can outperform ↑↑ flexible models b/c they avoid overfitting

SUPERVISED VS UNSUPERVISED LEARNING

- Supervised learning: For each obs. of input(s) we have an associated output
  - ↳ Goal: fit model that relates inputs to outputs ⇒ aim to accurately pred. future vals. of output OR better understand relationship btwn. input & output
  - Δ LR, GAMs, Splines, RF, NN, SVMs

- Unsupervised learning: for every obs. $i=1,\dots,n$, we see a vec. of measurements x_i but NO response y_i

↳ Can't fit supervised models w/o y

↳ What can we even do in an unsupervised learning situation?

- Look for relationships btwn. vars. or obs.
- Common tool: $\underbrace{\text{Cluster Analysis}}$

  find, on basis of $x_1, \dots, x_n$, whether obs. fall into DISTINCT groups

Δ (Pg. 25) Mkt. segmentation → find groups & see if they have diff spend habits

- For 2D probs, can sometimes just use eyes ⇒ for $\mathbb{R}^n$, $n>2$ probs, much harder to plot all obs. across all vars.

↳ Clustering becomes much more useful

- Sometimes have y only for fraction of n obs ⟶ semi-supervised learning

REGRESSION VS CLASSIFICATION

- Vars are either: (1) Quantitative: Have numerical vals.

  Δ Age, height

  (2) Qualitative: Have vals. in 1 of k diff classes / categories

  Δ Marital status, brand

- Regression Problems = Quantitative Response
- Classification Problems = Qualitative Response

Not always this clean-cut!

Δ Use LR for quant. & Log. R for qual.

↳ KNN & boosting (along w/ some others) can be used for BOTH

★ Regression vs Classification depends on OUTPUT type NOT input ⇒ can use quant. OR qual. for both types of models (as long as qual. vars. are coded properly) ★