Lec 14

Friday, February 10, 2023 1:31 PM

Recap

- dimensionality reduction (PCA), clustering (k-means)
- PCA : $f(x) = \mu + Vx$ we estimate μ and V from data.
  - $\hat{\mu} = \bar{x}_i$ V = Eigen vector matrix of covariance matrix X
  - $X_{N \times d} \rightarrow Y_{N \times q}$ (should be a digonal covarince matrix)
- clustering using k-means

  assignment ↓↑ template update → set of templates, group assignments

Modeling : Selection, fitting, and validation

selection: linear (vs) exponential mode

$y = aT + b$ (M_1)

$= b\exp(-aT)$ (M_2)

} prior knowledge, or by visual inspection

- Space filling design spanning T and pH and computed reaction yield
- 1. look at the data to identify trends that can help model selection
  2. pick a model from the class of selected models
     - hypothesis class

       eg: $r = aT + b$

       $= aT + bT^2 + c$

       ⋮

       } polynomials

       pick the right polynomial

     - Bayesian model selection.

       two or more models: $M_1, M_2, \ldots$

       $p((x = (T, pH)), y = r \mid M_1)$

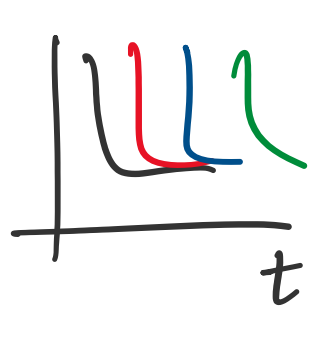
       M_2

       ⋮

       Bayes factor / simply look at likelihoods

- fitting / training : find parameters of our parametric models.
  - by some form of minimization of a loss function

    $$l = \sum_i | y_i - f(x_i) |^2$$

  - model  P_1, P_2, P_3, P_4

    $\Rightarrow b\exp(-at)$

    (a, b)?

    ↓ $f(P) = (a, b)$

    P

    $l = [(a, b) - f(P)]$

    eg: thermal history of 3D printed materials

  - ML and other advance techniques allows us to use arbitrary loss function, automatic diff to perform stochastic gradient descent.

- data driven modeling : universal function approximator

  $f(x) = W^T \sigma(x) + b$ — any continuous function using the deep network.

  no free lunch — not every model will be able to explain the data you have.

  - train data, validation data, test data.

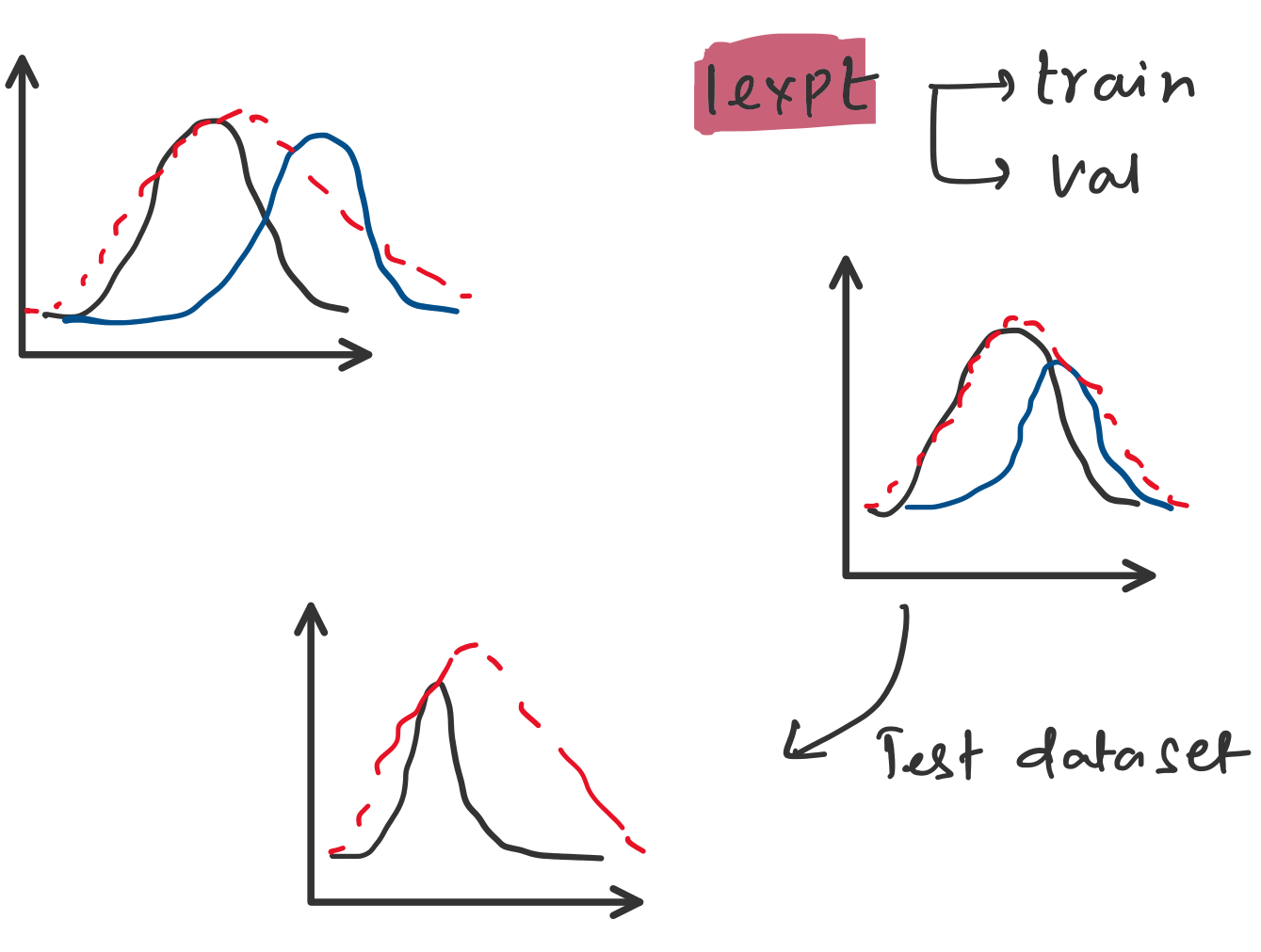
- data using any of DOE methods we disussed

  ↙ train (DOE) ↓ Val (DOE) ↘ test (DOE)

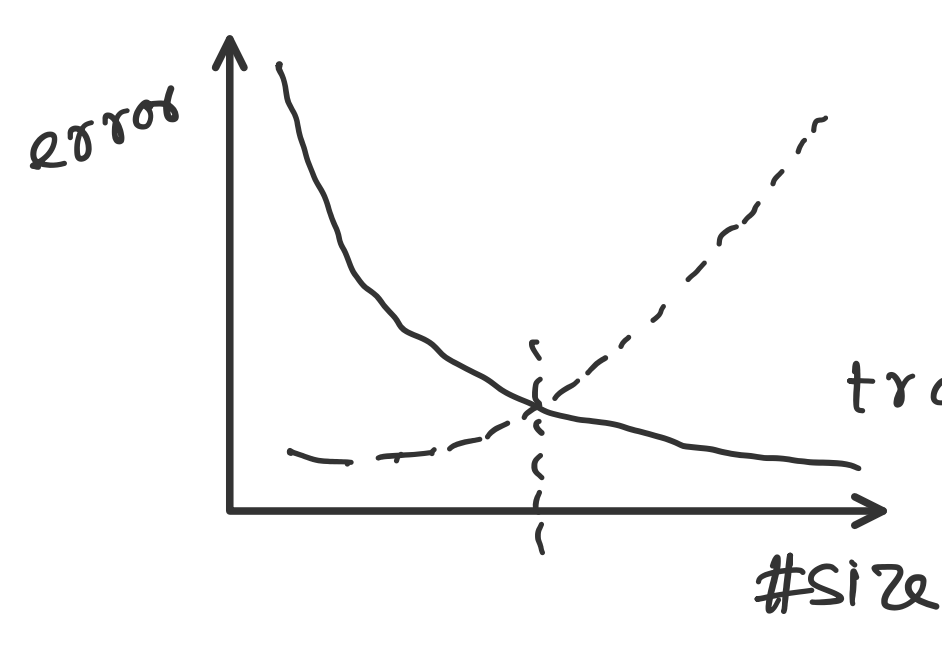
  ← 1 expt (high accuracy)

  ← 3 expt (most usefulness)

  every time you run an expt → "distribution shift"



- "sweet spot" high training accuracy, high validation accuracy.



training, val error as minimal as possible.

- $r \rightarrow (T, pH)$

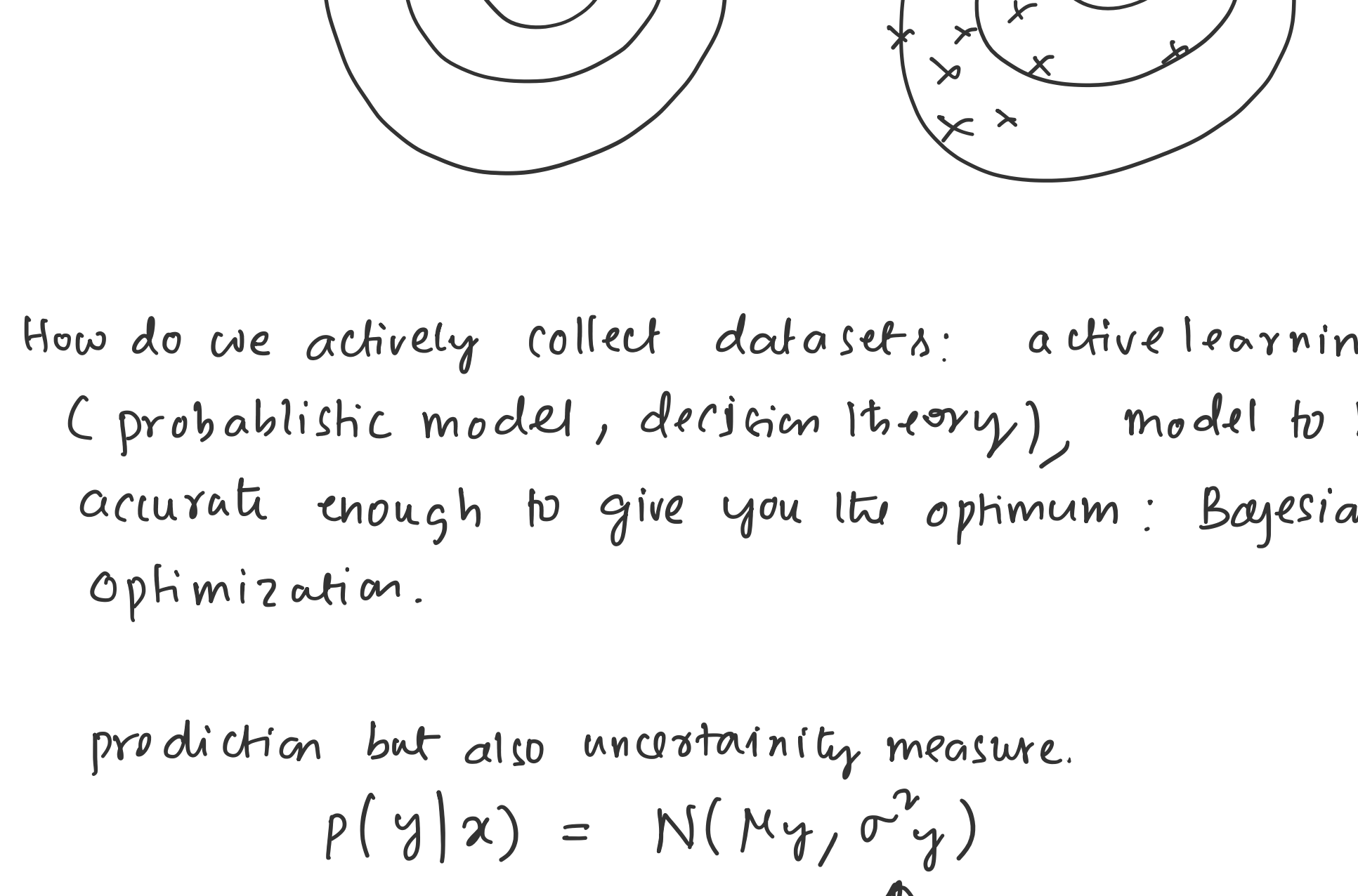
  $$p(T, pH) = \mu_r + \varepsilon \sim N(0, \sigma^2)$$

  (probabilistic model)

  bias, variance and noise — from the data itself; represent your data errors in expt that are not captured.

  variance : "overspecied" your model is to the data

  bias : particular parameters that were selected as solution.



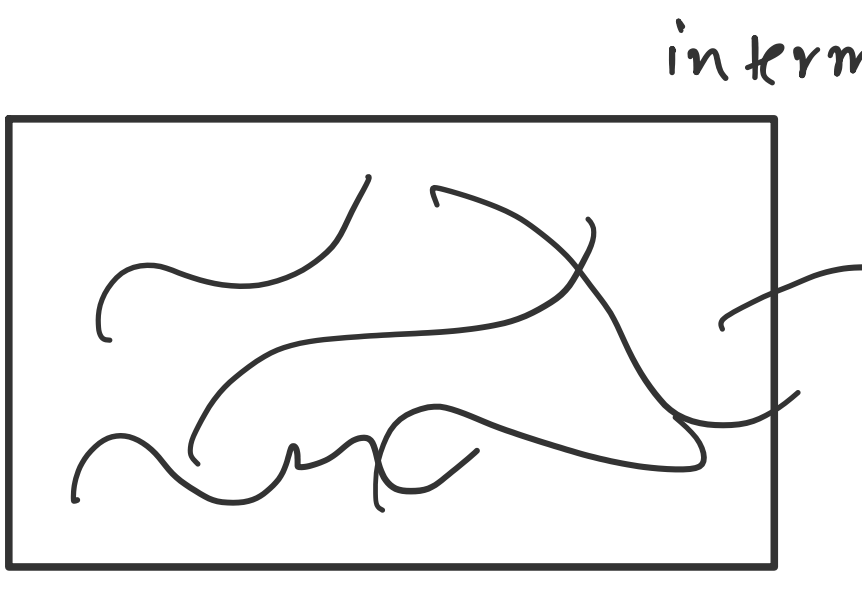
- How do we actively collect datasets: active learning (probablistic model, decision theory), model to be accurate enough to give you the optimum : Bayesian optimization.

- prodiction but also uncertainity measure.

  $$p(y \mid x) = N(\mu_y, \sigma_y^2)$$

  ↑ decsion models to suggest wher to sample next.

  Gaussian process : class of functions that have a common feature in terms continuity



$P(f) = 0.5$

$p(y \mid f(x))$ is high.

⇒ vector based approach → likelihoods (MLE, MAP) fit GPs