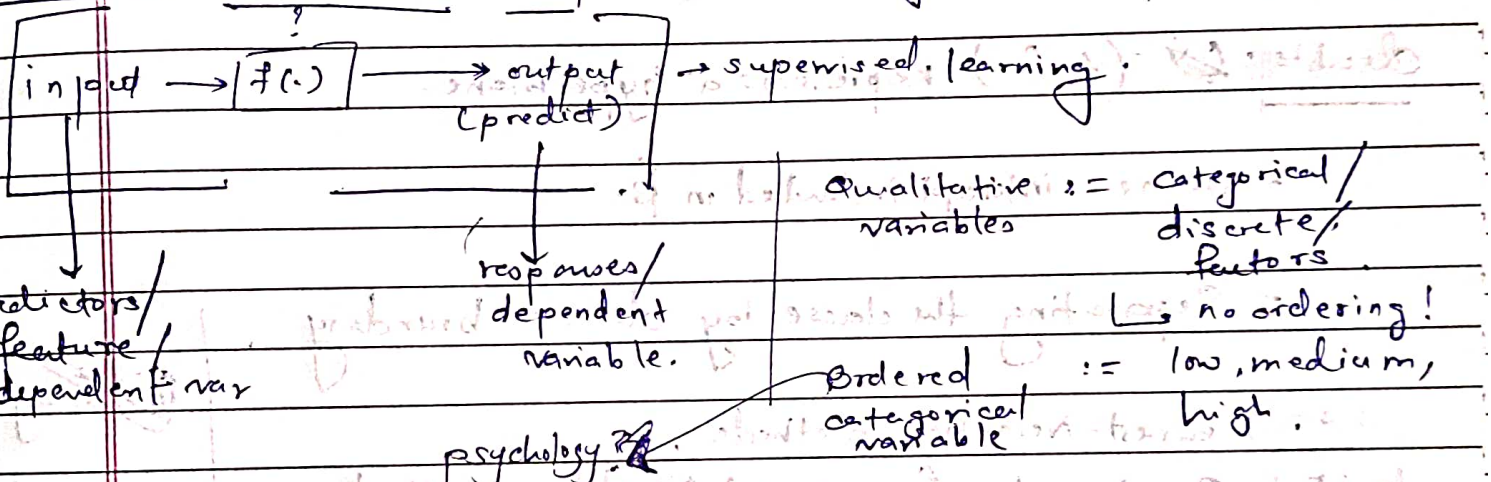
Elements of Statistical Learning:

Chapter 1: Introduction:

Regression & classification → supervised.
Gene-expression: clustering together!

Chapter 2: Overview of Supervised Learning

regression (prediction of quantitative)
classification (prediction of quantitative).



- targets: survived/died → 1/0 : one hot encoding.
- K level qualitative variable → K binary (1/0) var/bits := this are called dumy variables.
- data to construct prediction rules: training data.

doubt: what is it meant by stability of model?

* Diff in tabular form for Least Squares & Nearest Neighbors.
  ↳ take a data & apply both to same! → Task!

* The linear model fit by least squares:
  → huge assumptions about the structure
  → stable predictions.
  → possibly inaccurate predictions.
* K-nearest-neighbor prediction rule:
  → very mild structural assumptions.
  → unstable predictions.
  → often accurate predictions.

$x^T = (x_1, x_2, \ldots, x_p)$

$$\hat{y} = \hat{\beta}_0 + \sum_{j=1}^{p} x_j \hat{\beta}_j \longrightarrow \boxed{\hat{y} = x'\hat{\beta}}$$

bias.

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & & & \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

to include the bias.

doubts: $(X, \hat{y})$ represents a hyperplane.

- assume: intercept included in $\hat{\beta}$.

- Separating the classes by decision boundary.

2.3.2. Nearest-Neighbor methods:

doubt: Pseudo code for Nearest Neighbor methods.

To do: Apply clustering (KMEANS) on iris.
What exactly is the math behind it?

2.3.3 Read again.

- classifies a data point, based on how its neighbours are classified.
- k := parameter that refers to the number of nearest neighbours to include.
- finding value of k: parameter tuning.
- Nearest Neighbor for classificatn & regression → are different?
- Nearest neighbor methods uses those observations in the training set T closest in input space to x to form $\hat{y}$. Specifically the k nearest neighbor fit for $\hat{y}$ is defined as:

$$\hat{y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i$$

Notes: $N_k(x)$: neighborhood of x, defined by closest k many points. where 'closeness' is distance metric.