

January 19, 2023

DSML: CC Fundamentals.

Feature Engineering - 2.

Recap:

- (a) Hypothesis testing - χ , p-value, Null, Alternate.
- (b) Z-test } 1 sample, 2 sample, independent, paired,
- (c) t-test } left-tailed, right-tailed, 2-tailed.
- (d) R-S test
- (e) ANOVA
- (f) Kruskal
- (g) Pearsonr
- (h) χ^2 test

Class begins @ 9:05 p.m.



Recap : Data Analysis and Feature Engineering.

Solutions to Prediction problems.

1] Human driven: Statistics + Intuition.

→ 2] Machine driven: Optimization & Geometry.

→ 3] Hybrid: Statistics + Optimization.

loan data.

Key observations:

1] Missing data.

2] Income vs. loan status.

3] Total income vs. loan status.

Lessons learnt.

① Exploratory data analysis.

② Test relevance of features.

③ Creation of new features.

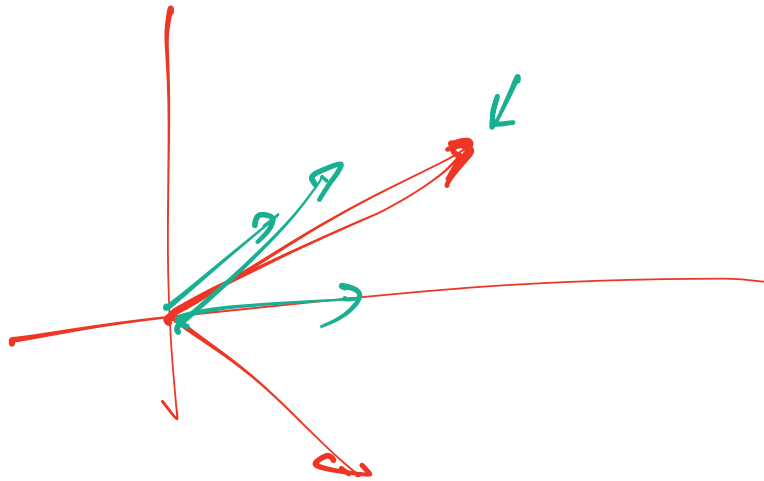
④ Feature transformation.

Today:

(a) Missing data.

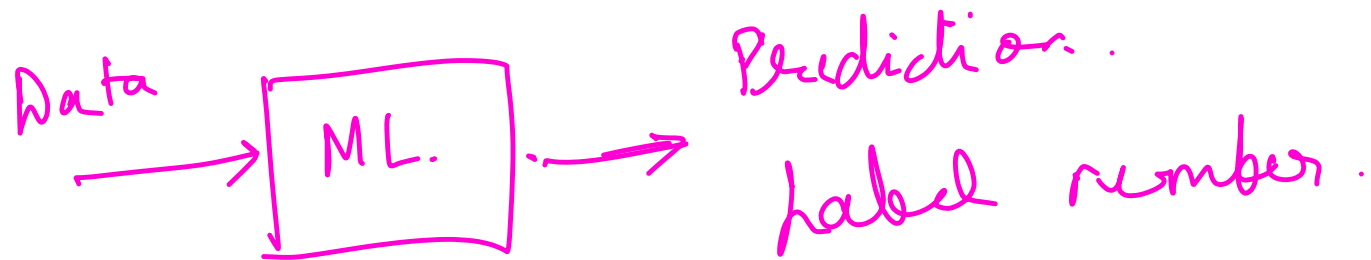
(b) Feature engineering: Creating use ful features.

(c) Encoding categorical variables.



K - nearest neighbors
input.

Prediction problem.



Log Lat Human. ←

→ [0. 1 0] ←

→ [0.1 0.8 0.1]

$P[\text{Loan_Status} = 'Y' \mid \text{Property Area} = \text{"Urban"}]$