AI - ENGINEERING BLUE-PRINT

- Dream in years
- Plan in months
- Evaluate in weeks
- Ship Daily

- Prototype for 1X
- Build for 10X
- Engineer for 100X

- What's required to cut the time-line in half
- What needs to be done to double the impact.

Video - series - 1 :

- Machine-learning Basics (Supervised Learning)

$y = f(x)$ ⟶ "its a form of equation

y ⟶ output
$f()$ ⟶ function form
x ⟶ Input

2) y - output ⟶ stock ⟶ millions
accuracy - out of sample

time-series - continuous → Mean Squared Error (MSE) $= \frac{1}{n}\Sigma(y_i - \hat{y}_i)^2$

Classification ↙ MAE (Mean absolute Error) $= \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$

Accuracy = $\frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$

RMSE = Root Mean Squared Error $= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

Precision = $\frac{\text{True Positive}}{\text{true positive + false positive}}$

$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$

Recall = $\frac{\text{True Negative}}{\text{true positives + false negatives}}$

MAPE = Mean Absolute percentage Error (MAPE) $= \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$

F1-score = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}}$

ROC - AUC (Reciever Operating Characteristic - Area Under Curve → Measures the ability of classifier to distinguish b/w classes. A higher AUC indicates better model performance.

- Logarithmic Loss (log loss): → 0 and 1 performance of classification model whose output → 0 & 1

$$\text{LogLoss} = \frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$$

Classify Cats
1. Collect Images (Source)
2. Label Images → cat / non-cat
3. Pre-process Image
   1. Resize Images (128 x 128)
   2. Normalization (scaling pixel values between 0 & 1 dividing by 255)
4. • CNN's
   • Pre-trained Models like VGG16, ResNet, or MobileNet → pre-trained on ImageNet → can be fine tuned to specific task using (transfer learning)
5. Build and Compile the model
6. Train the model
7. Test the model

independent (X) — (Y) dependent

education → salary.

salary = β_1 Education + β_0

β_1 ↔ best estimate.

2) $f(x)$

1) Defining variables

(X) : Education level → • years of education
• highest degree

(Y) : Salary (continuous variable)

2) Formulate the regression model.

$Y = \beta_0 + \beta_1 X + \epsilon$

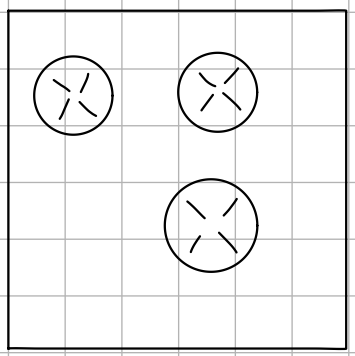
↓

If family-background comes into picture

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

Y = dependent variable (salary)
X = independent var (education)
β_0 = intercept (salary when edu is zero)
β_1 = slope-coefficient (the change in salary for each additional unit of edu
ϵ → error-term

3) X-Inputs Optimization :- are central to ML, as they involve finding the best parameters for a model to minimize (or maximize) a certain objective function. → error or loss of model, and optimizing it means improving the model's performance on a given task



Case (1) :- Handing out coupons leads to ↑ customer spending or wrong ?)

→ wrong

(benchmark)

RCT :- Randomize controlled trail bias spend

(FOMO)

coupon → cost

• customer who get coupon are likely to spend anyway

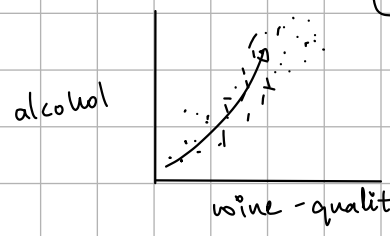
Fb Y Tv P

$y =$ (|||||||) $(1-9)$

sign p-ve → 1/20

1-8 wine-quality → alcohol, texture, sulpher-content, nitrogen

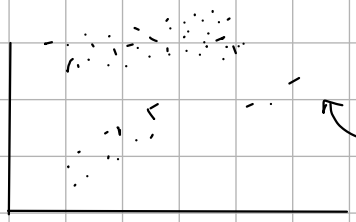
(9)



(Wrong!)

Bias → refers to a systematic error that occurs when a model consistenly makes error

Unbiased → in a statistical sense, makes predictions that, on avg, are correct. It does not overpredict or underpredict the true values across many data samples.



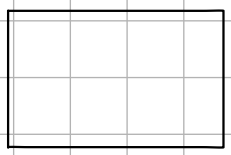
Here lies the dragon in statistics

(Case iii)

risk of relapse

X wrong

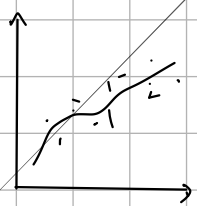
Ideal-scenario

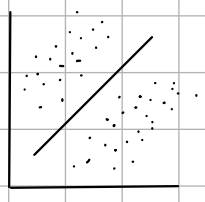


(Model) – NN – 99%

$y = f(x)$

Discriminative Models



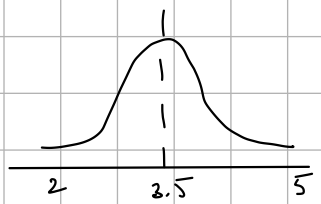


frequentist-setting $y = \beta_0 + \beta_1 x + \varepsilon$

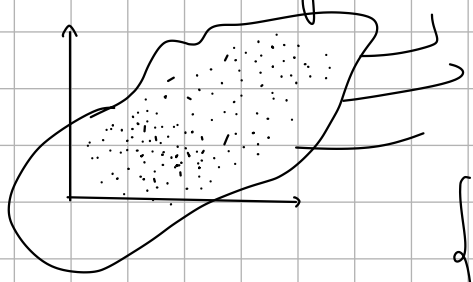
Probabilistic setting

$P(y|x)$

$\beta \rightarrow$



Generative Modeling



∞

figuring out data

- GMM
- GAN
- HMM

ML - pipeline → blue-print

1) What is the business usecase ?

→ (USA) → rescue - equipment → Hurricane

→ path of hurricane

we can tell the rescue equipment

8 → 3 samples

10000 → 3997 → 1-7

2) Data →
   i) Infinite data would I be able to solve the problem
   ii) Given that I will have enough variation of the data

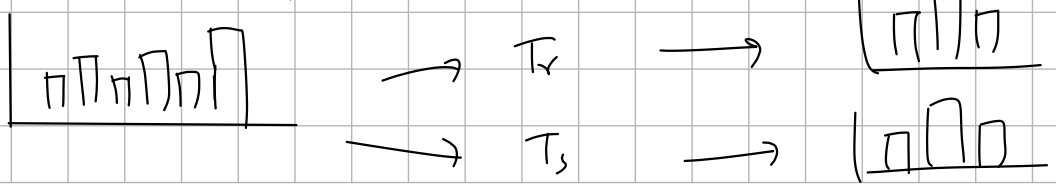
3) Define Error - metric
   - TS → TS → MAPE, MSE, RMSE
   - classification → Accuracy, precision, recall, ... - f2-score

Aim of modelling process to min OOS error.

T_r T_s

'generalization error'

- Balanced Split



1) Proper Data splitting :- 1) Stratified Sampling (Classification)

↓ ensure distribution of classes in the training and test sets is similar to that in the original dataset.

eg:- emails → spam
→ not spam

} stratify = y ensures that we have same proportion of each classes as the original dataset.

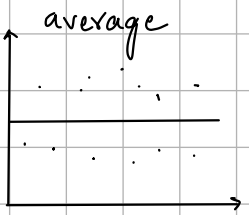
Regression example : Random sampling → output variable is continuous
: Cross-validation → assessing models performance on different subsets of the data. which helps reduce generalization error.

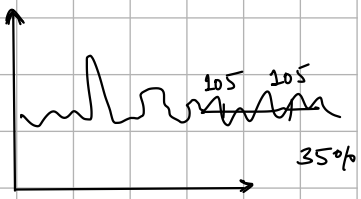
3) Determine the task → whether it is regression classification

6) Setting the base-line

(no-covid) 95 % (covid) 5 %

(95% baseline-accuracy)



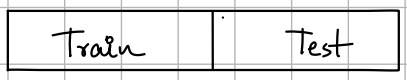


37% not beating your baseline

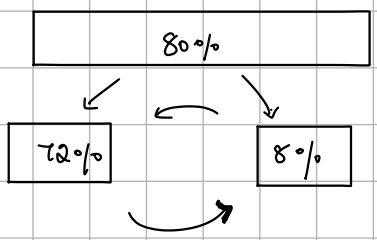
{95% accuracy} (17% MAPE) min-accuracy

↳ if the client doesn't provide baseline, we set up our own baseline

(cold-storage)



7) Apply models on train and cv-set.



* Logeistic regression
* Random Forest
* SVM
* Gradient Boosting Method

Cross-validation (CV)

8) Prediction interval }
Confidence interval }

---

1) Books to read → * Rob J Hyndman and George Athanasopoulos

Forecasting Principles and Practice (2nd ed)