ML- U3

Supervised Learning- Regression

classmate Date ____ Page ____

Bias-

- Bias → • phenomenon that skews result of an algorithm in favor or against an idea
  - systematic error in ML models → incorrect assumptions.
  - Diff b/w average prediction of our model & correct value
  - Bias & variance → components of reducible error.

- Low Bias → fewer assumptions → form of target fuction → by model
- High Bias → more assumptions → unable to capture imp features of dataset → can't perform well on new data.
- examples → Low Bias → decision trees, k-nearest nighbours & SVM.
- examples → High Bias → linear discriminant analysis & logistic regn.

* Ways to Reduce High Bias-

- Increase input features as model is underfitted.
- Decrease regularization time period
- Use more complicate modes → that have polynomial features
- Increase training data.
- Increase regularization term.

classmate
Date
Page

Variance -

- changes in model when using different portions of training dataset.
- how much random variable differs from its expected value.
- Based on single training set.
- measures inconsistency of diff predictions using diff training sets.
- Model → high variance → low bias & viceversa.
- comes with highly complex model with lot of features.
- contributes to flexibility of model.

high variance signs → • Noise in dataset
• overfitting
• complexity
• Forcing data points together.

- Low variance → small/low variation in prediction of target functions.
- High variance → large variation in prediction of target fuctions.

To Reduce High Variance

- Reduce input features or no. of parameters as a model is overfitted.
- Do not use a much complex model
- Increase training data.
- Increase the regularization term.

Bias & Variance trade-off —

- Tradeoff b/w bias & variance is inversely proportional to variance & viceversa
- tackle by either increasing a) complexity of model
  b) Training dataset.

Overfitting.

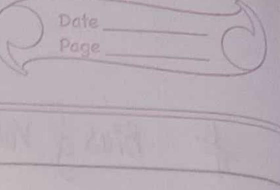
- statistical model fits against its matching data.
- Algo/ model can't perform well on unseen data.
- low bias & high variance.
- ~~unable to~~ match input data to target data. / non-esistent
- model → complex enought → match all datapoint & performs well
- Reasons → • noisy data
  - training data is too small
  - large no. of features.

To avoid Overfitting

- Cross validation
- Train with more data.
- Removing features
- Early stopping the training
- Regularization
- ensembling.

Underfitting -

- Data model is unable to capture relationship b/w input & output.
- occurs when model is too simple.
- High bias low variance.
- larger quantity of features.
- less data → for model.



To avoid underfitting.

→ • growing educational time of model
→ • By increasing the wide variety of functions.

Regression -

- functional relationship between two or more correlated variables that often empirically determined from data & used to predict values of one variable given value of others.

- variable → influence on output → known as explanator, indepd. input or predictor variable.

- variable → outcome depends on other variable → dependant variable
- dataset in which target values are known.

(correlation b/w depedant & indepent variables).

desired ouput → one or more continous input → regression

Linear Regression –

- ml algo → supervised learning
- predict dependent variable based on independant.
- finding straight line that best approximates set of pointer on graph.

- simple linear regression → only 1 independant variable
- mutiple linear regression → more than 1 independant variable

  x → independant variable

  y → ~~dependant variable~~ Output.

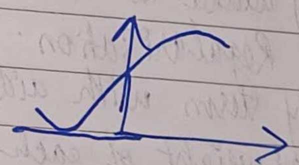
- $y = \beta_0 + \beta_1 + B_2 x_2 + B_3 x_3 + \cdots B_n x_n + \epsilon$

Disadv → • outliers affect.

• oversimplifies real world problems.

Use cases → • Demand Forecasting • Healthcare • Other predictions.

Logistic Regression.



- used to model & estimate probablity of an event
- estimate problem probablity of an event occuring → value (indepedent)
- Is a curve.
- sigmoid curve or S-curve in short.
- outcome data points → 0 or 1.
- 0 → totally uncertain 1 → certain
- solve clarification issues.
- predictive evaluation algo.
- predict likelihood of categorial dependant variable.

Type
- Binary → 0/1, pass/fail
- Multi → Cats, dogs, lion
- ordinal → low, medium & high

$$f(x) = \frac{1}{1+e^{-x}}$$

- Lasso Regression -

- lessen complexity of version
- similar to ridge regrⁿ except that penalty time includes only absolute weight instead of rectangular of weights.
- L1 Regularisation
- avoids overfitting of training data & useful for feature selection
- shrinkage method. (least absolute selection & shrinkage operator)
- overcome disadv. of Ridge by not punishing high values of coefficient $\beta \rightarrow$ actually setting them to ze if not relevant.

$$L(x,y) = Min\left(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + \lambda \sum_{i=1}^{n} |w_i|\right)$$

* Ridge Regression -

- small quantity bias delivered → higher long time predictions-
- quantity added to bias → ridge regression penalty
- L2 Regularisation.
- penalty term with aid of multiply with lamda to the squared weight of each individual features.

$$L(x,y) = Min\left(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + \lambda \sum_{j=1}^{n} (w_i)^2\right)$$

- Reduce complexity of verse.
- Clears up trouble if we have greater parameter than samples

↑ α ie λ → more coeffient become robust to collinearity.

· Gradient Descent Algo- (finds best fit line for giving training datset)

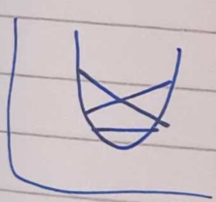
· optimization algo → minimize cost function.
· locate least possible value → fulfill a given fee fuction.

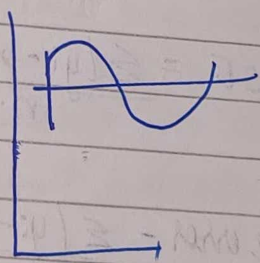
Types

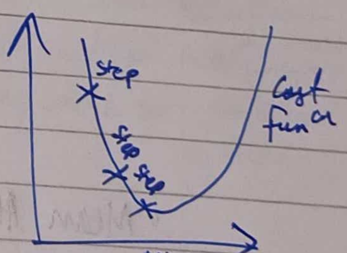
· Batch → process all training ex. for each iteration.
· stochastic → Process 1 traing example per iteration.
· Mini Batch → More examples out of total m training examples are processed per iteration. Faster than Batch

funcⁿ requirement

· Differential - $f'(x)$ exists for every x in domain
· convex - for an univariate funcⁿ line segment connecting two funⁿ points lays on or above its curve.


Convex


Non Convex.



~~Features~~

Limitations → · relatively slow close to min.
· for poorly conditioned convex problems gradient points orthogonally to shortest direction to min. speed

Evaluation Metrics / Cost function -

- establish events or set of values to represent loss or penalty incurred in gaining something.
- Aka loss or error function.
- The objective f's that need to be maximized are called reward/point/utility functions.

- Mean Error (ME) - mean of diff b/w actual & predicted value

$$ME = \frac{\Sigma(y_i - x_i)}{n}.$$

- Mean squared error - diff b/w authetic & expected values extracted by squared the average distinction.

$$MSE = \frac{\Sigma(y_i - x_i)^2}{n}.$$

- Mean Absolute error - $\frac{\Sigma(y_i - z_i)}{n}$ → .

- RMSE Root Mean Squared error — P

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\Sigma(y_i - x_i)^2}{n}}.$$

$$R^2 = \frac{\text{sum of squares due to regression}}{\text{total sum of squares.}}$$

↙

shows how well data fits regression model.
↑R^2 → Better is model.