Lecture - HTuning, CV, k-fold, log Reg

https://colab.research.google.com/drive/1c21GqeA5S0do0JZ2H6_Uc0DRi8WF1SgD?usp=sharing

Agenda

→ Hyperparameter Tuning
→ Cross Validation } Generic
→ K-fold Validation }
→ Intro to Logistic Regression

Hyperparameter Tuning

Parameters

$w_0, w_1, w_2 \ldots w_d$ - Learnt through training

Hyper-parameters

λ, α, Degree of polynomial - set by ML/DS Eng

Q Regularization Rate (λ)

$L = MSE + \lambda \sum_{i}^{d} w^2$ | | ( )²

L1 L2

Case 1) $\lambda = 0$ - NO Regularisation → May overfitted

Case 2) $\lambda = 10$

Case 3) $\lambda = 100$

Case 4) $\lambda = 100000 (\infty)$ ↑ Regularisation → Underfitting

How to choose opt. value of λ ?

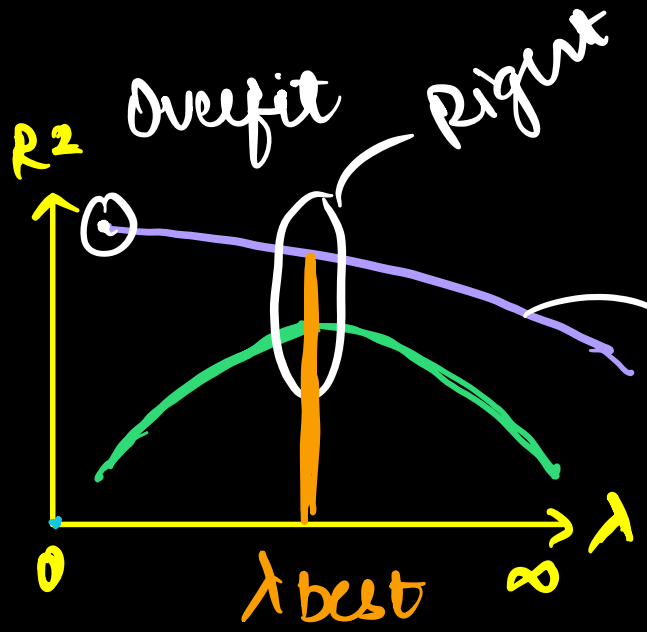
$\lambda = [0, \infty)$; $\lambda = [0, 0.001, 0.01, 0.1, 1, 10, 100]$

Train diff. models with ↑ diff λ

CHOOSE THE MODEL WITH BEST PERFORMANCE

TRAIN DATA — favour overfitted model ✗

TEST DATA ✓



As ↑λ, → Underfitting

Cross Validation

Chose λ using test data.

1) Train models with diff values of λ
2) Chose the model w ↑ R^2 on test data.

↓

Data Leakage

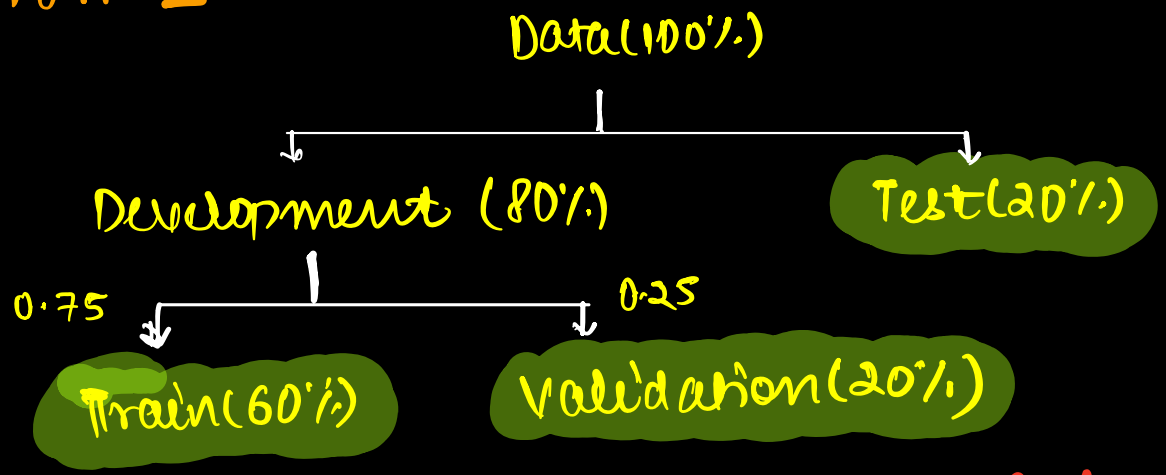
Is my TEST DATA really kept unseen? NO

⇒ Best λ may overfit to test data.

loss of generality

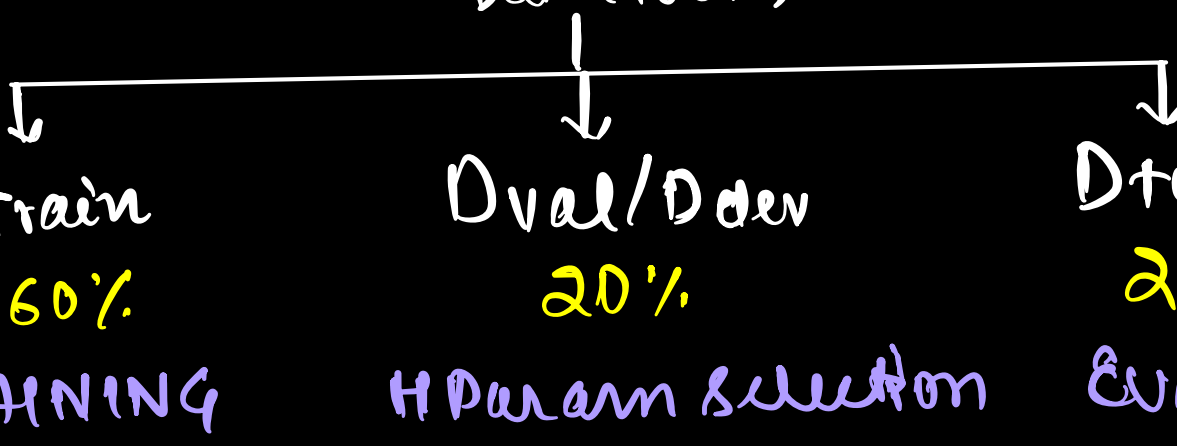
(CAN'T COMMENT ON Generalization based on test data.

Solution-1



My personal style (used by most of best Researchers)

Popular



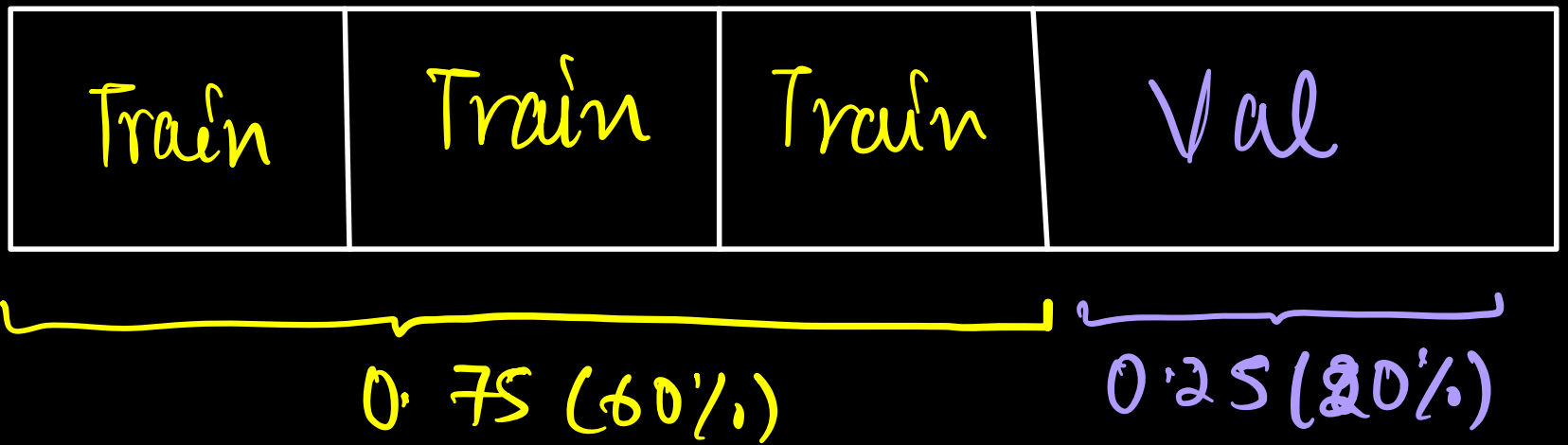
Code - Cross Validation - Colab.

Problems with Cross Validation

1. Decrease in size of train 80% ⟶ 60%
2. We may end OVERFITTING to val data

Solution

80% Dev Data



Lets do the same k-times

K-fold cross validation :

	1	2	3	...	K		
Fold 1	Train	Train	Train	Train	Val	CV1	λ_1
Fold 2	Train	Tr	Tr	Val	Tr	CV2	λ_2
Fold 3	Tr	Tr	Val	Tr	Tr	CV3	λ_3
:	Tr	Val	Tr	Tr	Tr		
Fold K	Val	Tr	Tr	Tr	Tr	CVK	λ_K

$$\text{Best } \lambda = \frac{\lambda_1 + \lambda_2 + \lambda_3 + \ldots \lambda_K}{K} = \lambda_{avg}$$

↑ Final λ

Typically $k = 5, 7, 10$,

V.V. less data - m-fold CV

Typically

Data - 100, 1L datapoints

60/20/20

Data > 1L

80/10/10

Data > 1M

90/5/5

Data - Big data

98/1/1

Lecture - Logistic Regression

https://colab.research.google.com/drive/1c21GqeA5S0do0JZ2H6_Uc0DRi8WF1SgD?usp=sharing

→ First classification model = LogReg ~~Regression~~

→ LinReg $\xrightarrow{t}$ LogReg

→ Geometric Interpretation

Notation

no. of samples - m
no. of features - d
no. of classes - n

first axis

X - (m, d) ↓
ith sample - $X[i]$, $X^{(i)}$
jth feature - $X[:, j]$, X_j

Output → y -

Nominal
One hot Encoding

Binary classification $(0, 1)$

$y \rightarrow (m, 1)$

Multiclass classif. n classes

$y \rightarrow (m, n)$

2 classes

1
0
1
0
1
0

mx1

3 classes

0	0	1
0	0	1
1	0	0
0	1	0
1	0	1

mx3

Class1 001
Class2 010
Class3 100

Binary

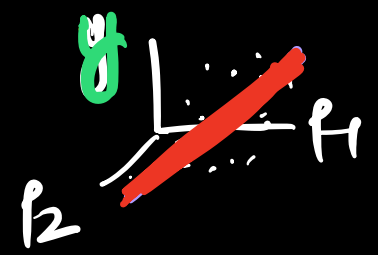
Multiclass

Parameters - $[w_0, w_1, w_2, w_3 - - w_d]$

Summary: LinReg $h: w^T x + w_0$ ★★

Linear model - Best line of fit passing through data

Assume d-features.



line $\in$ (d+1) dimensional space.

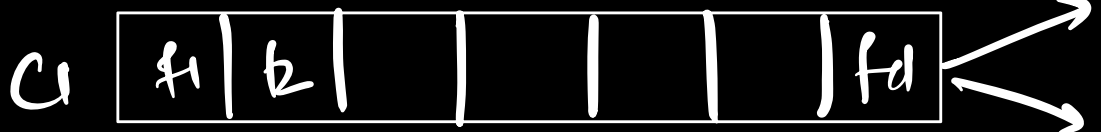
$$\text{loss} = \text{MSE} = \frac{1}{m}\sum_{i}^{m}(y^i - \hat{y}^i)^2$$

$y \in (-\infty, \infty)$, $y \in R$

Intro to Businesscase - Verizon/Airtel/VI

Customer Churn - leaving subscription.

Plan Rate, Network, statelcity, # disconnection
Customer Servie Feedback, # customer care.

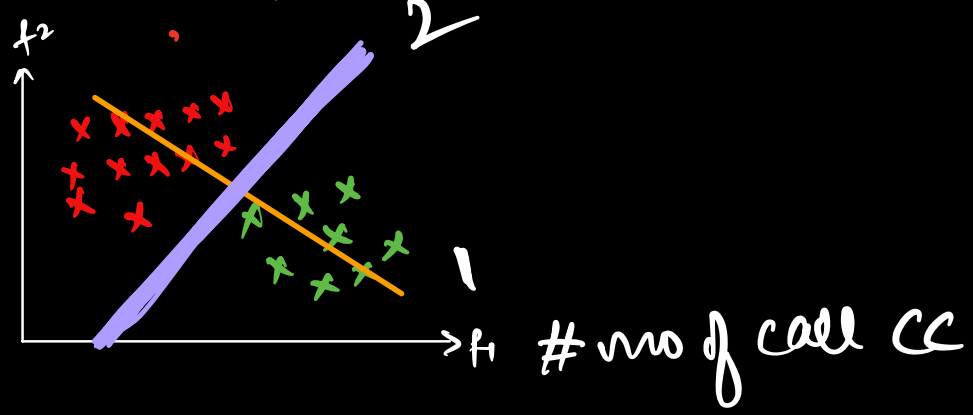


- Binary Classification - 1, 0 — not leave
- ↳ leave/churn
- Historical data with labels (1/0) - Supervised

- $y^i \notin R,\ y^i \in \{0, 1\}$

~~Class1 0 1~~ 0
~~Class2 1 0~~ 1

Charges for extra service



y is represent with colors not with axis.

1 line of best fit
d+1 space
one axis for y
$y = w^T x + w_0$

2 Decision Boundary
Best line of seperation
d space