Agenda

- Regularization
- Updated Loss Function
- Regularization parameter
- L1 Regularization
- Elasticnet Regularization
- Hyperparameter Tuning
- Cross Validation
- K-fold Cross Validation

Regularization

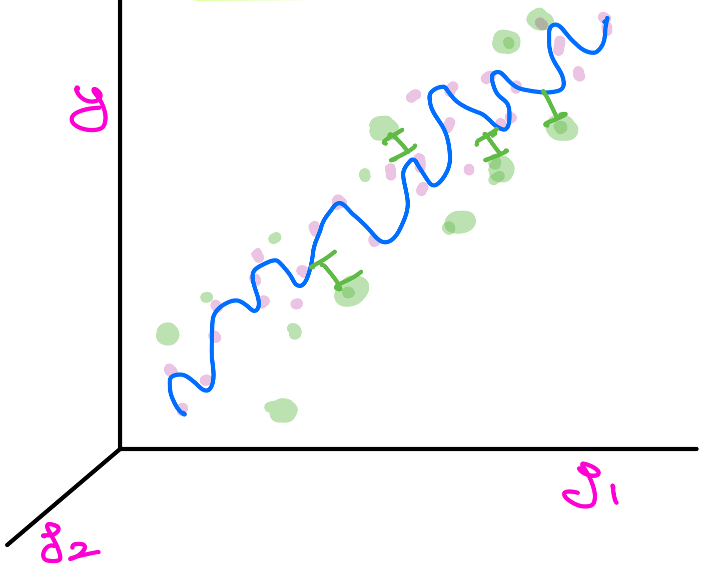
$0 \leftarrow 0.1$ (reduce Import)

$\uparrow$

10

$\uparrow$

$$\hat{y} = w_0 + w_1 f_1 + w_2 f_2 + w_3 f_1^2 + w_4 f_2^2 + w_5 f_1 * f_2$$



if model is overfitting? Yes

Complex Model
(features with high degree)

Fix
(reduce mag(weight) to reduce the Importance of features)

① idea 1 : Reduce features

(i) Reduce degree

(2) Manually Remove features by setting $w_j = 0$

② idea 2 : Modify the Loss function st $w_j \longrightarrow 0$

Loss : $$\min_{w_j} \underbrace{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{\text{Loss}} + \text{Regurization Term}$$

Regurization Term → minimizing mag(w_j)

Updated Loss Function

$$L \Rightarrow \min_{w_j} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{d}(w_j)^2$$

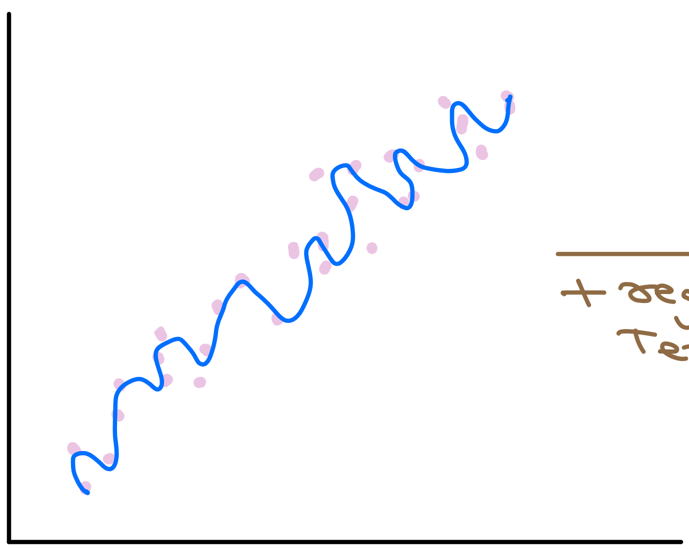
minimize mag $(w_j) \longrightarrow 0$

$$\frac{\partial L}{\partial w} \Rightarrow$$

$$\hat{y} \Rightarrow y_i - (\cancel{w^T x} + w_0)$$

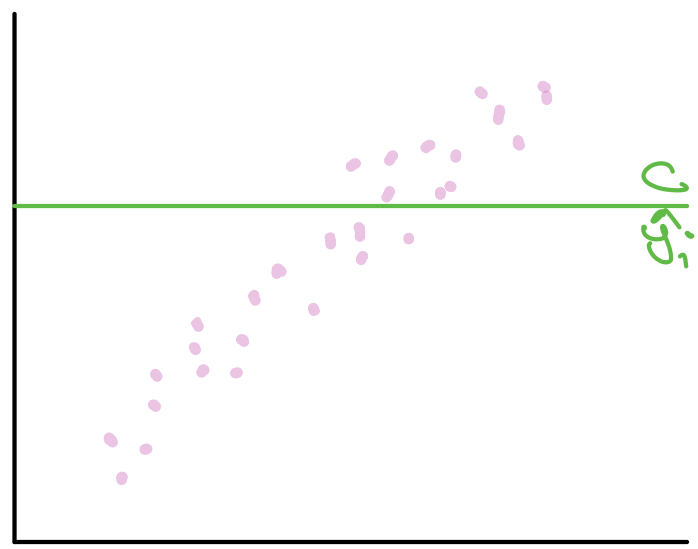
$$\cancel{w^T x} \to 0$$

$$\hat{y} \Rightarrow y_i - w_0 \Rightarrow c$$



Overfitted

$\longrightarrow$ + reg Term



Underfitted

Regularization Parameter

$$L \Rightarrow \min_{w_j} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{d}(w_j)^2$$

Loss Term (Variance)

Regularization Term (Bias)

λ → Hyperparameter → Act as Tuner To find Optional Balance

Bias and Variance Tradeoff

λ ↑ ($0 \to \infty$)

λ increase

Updates focus more on mag (w_j)

model Underfits

λ ↓ ($\infty \to 0$)

λ decrease

model Overfits

$$\boxed{\min_{w_j} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{d}(w_j)^2}$$

L2-Regularized MSE

↓

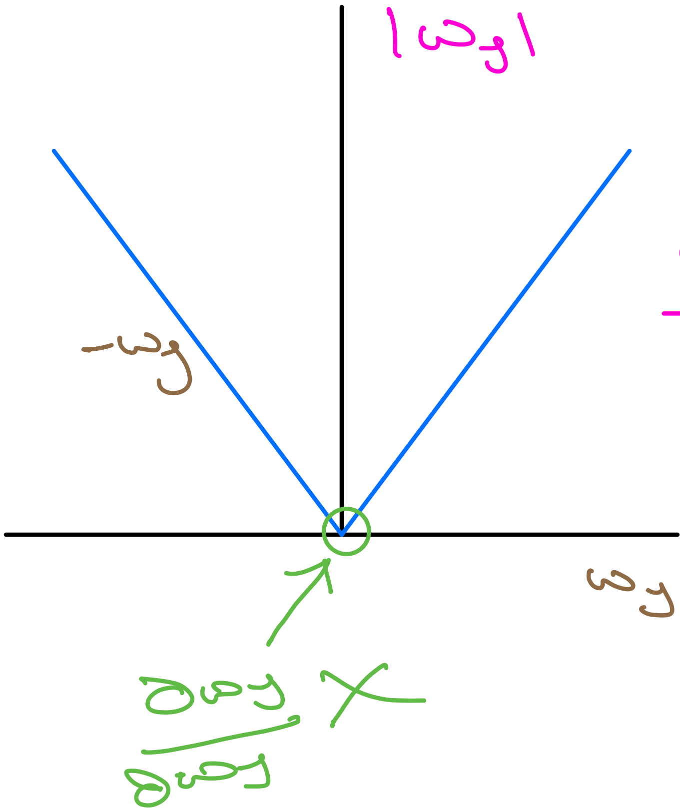
ridge

L1 Regularization

↓

Lasso

$$\min_{w_j, w_0} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{d} |w_j|$$



$$\frac{\partial |w_j|}{\partial w_j} \rightarrow \begin{cases} 1 : w_j > 0 \\ 0 : w_j = 0 \\ -1 : w_j < 0 \end{cases}$$

$$\frac{\partial w_j}{\partial w_j}$$

$$\frac{\partial (-w_j)}{\partial w_j}$$

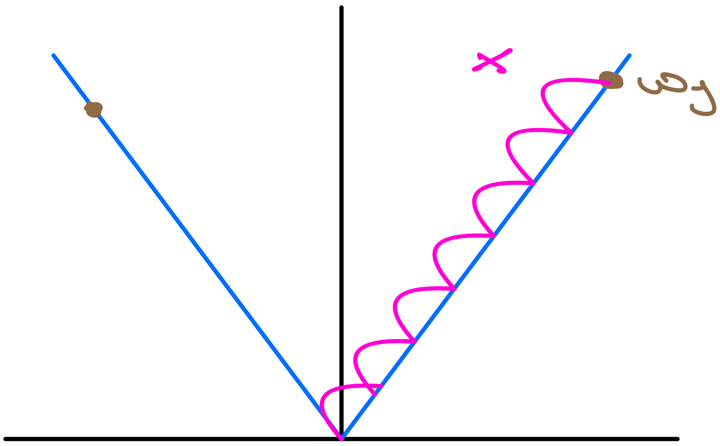
L1 Reg Term	L2 Reg Term
Can $w_j = 0$ (Can be used for auto feature Elim) ↓ Can create Sparse w_j	① Update to w_j will be smoother ② Will get impacted by outliers in weight

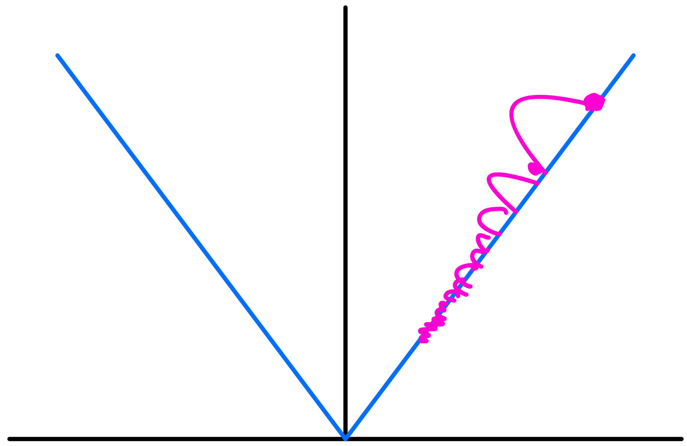
L1 Reg

$$\sum_{j=1}^{n} |w_j|$$

L2 Reg

$$\sum_{j=1}^{n} w_j^2$$





Update step :

$$w_j = w_j - \eta \frac{\partial L}{\partial w_j}$$

L1 Reg Term

$$\frac{\partial |w_j|}{\partial w_j} \quad \begin{matrix} +1 \\ 0 \\ -1 \end{matrix}$$

Reaches to 0 Quickly

L2 Reg Term

$$\frac{\partial (w_j)^2}{\partial w_j} \Rightarrow 2 w_j$$

Initially Fast but Slow down and never = 0

Elasticnet Regularization

$$\min_{w_0, w_j} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda_1 \sum_{i=1}^{d} |w_j| + \lambda_2 \sum_{i=1}^{d} (w_j)^2$$

Find Best Value Using Hyperparameter Tuning

L2 Regularization $\Rightarrow \lambda (w_j)^2$

$\rightarrow$ Balance use of all features

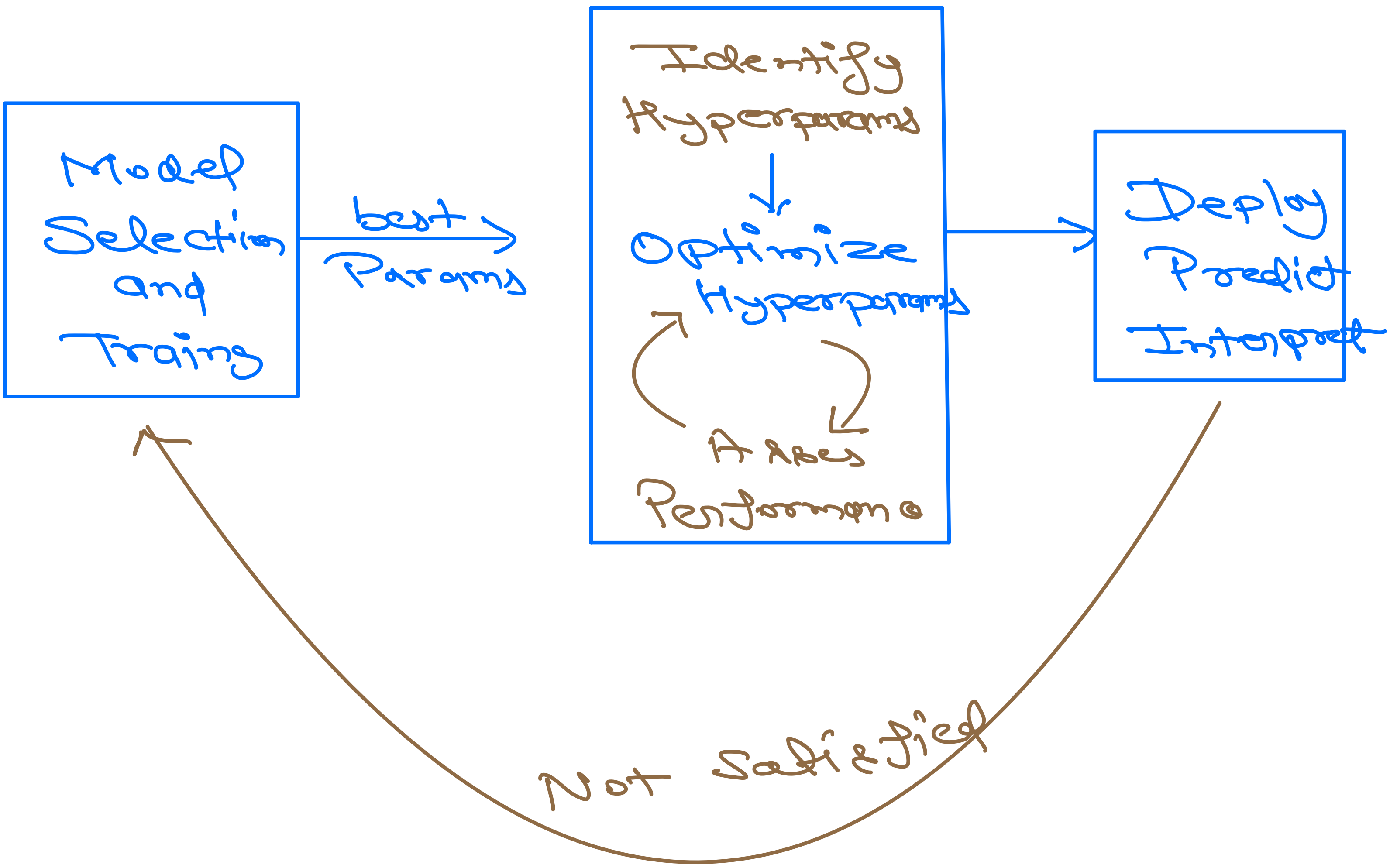
L1 Regularization $\Rightarrow \lambda |w_j|$

$\rightarrow$ Automatic feature Scaling

ElasticNet $\longrightarrow \lambda_1 L1 + \lambda_2 L2$

Hyperparameter

Parameter	vs Hyperparameter
Weights Learnt during model Training using the Train-set $\bar{w}$ and w_0	Value which we tune Manually Based on Experiment



Hyperparameter Tuning

Train-set	Test-set

0.80

0.81

Parameter

Optimization

Hyperparam

Tuning

0.81 0.83

R2_score

Accuracy

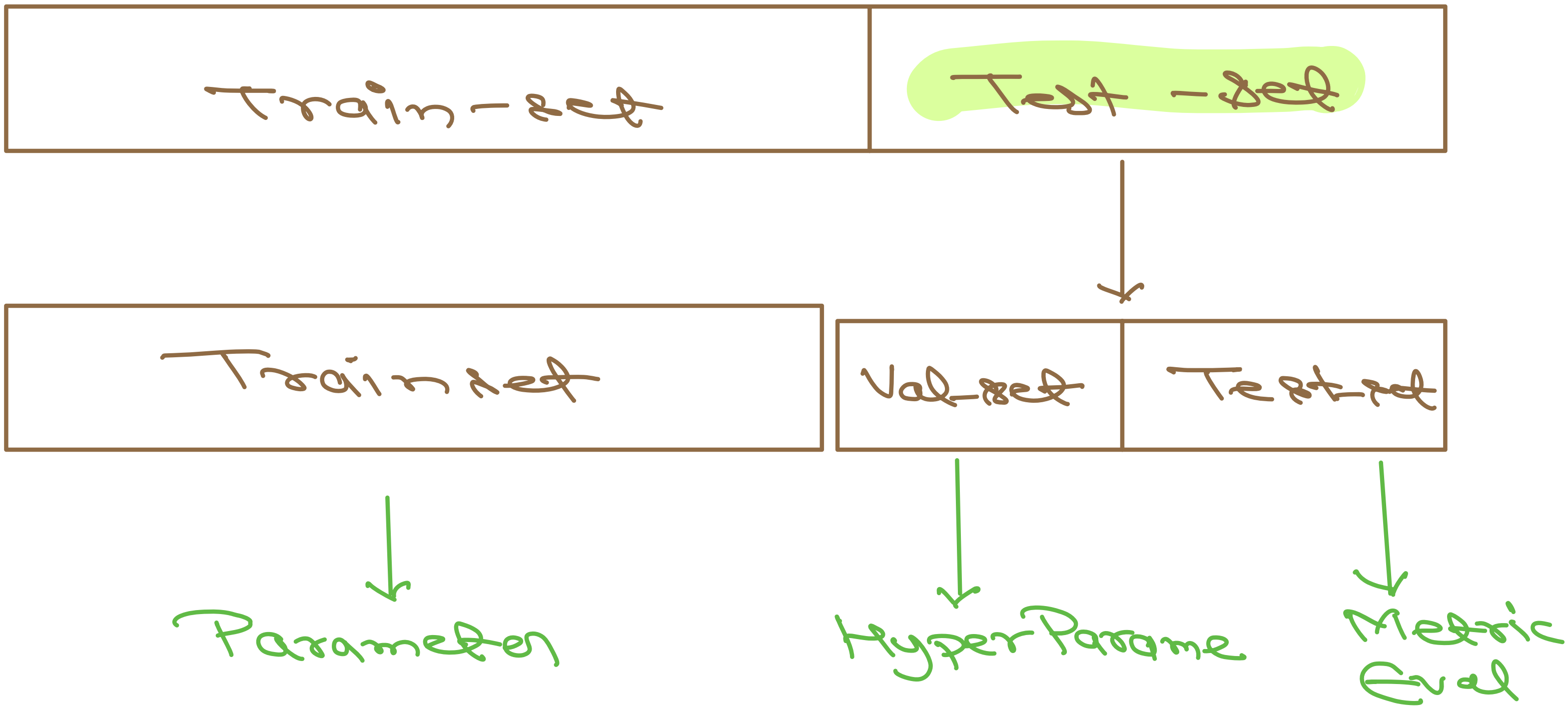
P, R, AUC

metric

Train-set

Test-set

Cross Validation



H.W: Explore

Grid Search
Random Search

Optional: Bayesian Search

K-fold Cross Validation

Full
Data
Set