* Regularization Techniques: Linear models.

- Also known as shrinkage methods.
- To build the perfect model we have to exercise permutation and combinations of independend features and target features.
- More the features or dimension, its harder to establish the relation, called Curse of Dimensionality.
- Also missing and outlier data creates issues.
- Due to these two reasons model fails to build perfect fit line. And generated best fit line is fitted between peaks and steeps.
- This happens due to high value of coefficents in cost function.

⊙ To fix this issue we introduce a penality or complexity or regularized term to cost function, which reduces the magnitude of coefficents without reducing the features.

$$\boxed{\text{Regularized cost function} = \text{Cost function} + \text{penality}}$$

- Their are several regularization techniques.

⊙ This cost function $J(\theta_0, \theta_n)$ can be any cost function.

- θ_n is slope, θ_0 are intercepts.
- θ_n can be $(\theta_1, \theta_2, \dots)$
- Equation of best fit line is like,

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 + x_2 + \dots + \theta_n x_n + \mu$$

[A] Ridge Regression [L2 Regularization]

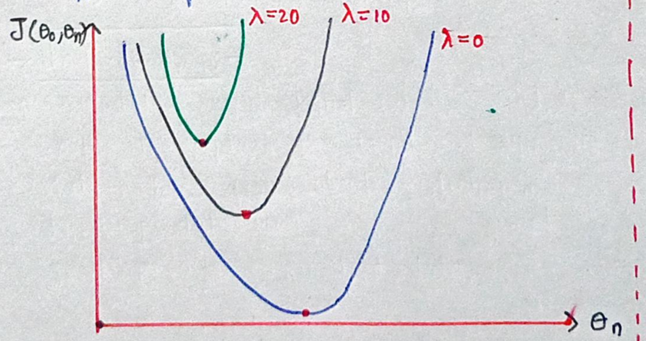
⊙ Coefficient of θ_n with high magnitude will generate the peaks and deeps of the graph, and to supress this we use penality factor (λ) which smoothens surface insted of an irregular graph

- Ridge regression tries to push coefficents of θ_n to zero, but not exact zero. As exact zero means feature gets deleted.
- This is L2 regularization since we add penality equivalent to square of magnitude of coefficents.
- As it tries to decrease magnitude of coefficents to zero it is good choice to be used when their is issue of outliers and we want to reduce overfitting.
- If $\lambda = 0$, its simple cost function.

$$\boxed{\text{Ridge Regression cost function} = \text{cost function} + \lambda \sum_{i=1}^{n} (\theta_n)^2}$$

⊙ λ is hyperparameter and we need to tune it.

- λ is inversly proportional to magnitude of coefficient of θ_n.



[B] Lasso Regression [L1 Regularization]

- Concept of working similar to Ridge.

⊙ But here magnitude of coefficent (θ_n) allowed to become zero while being minimized, this leads to elimination of feature ultimately leading to feature reduction and helps deciding which attribute good for relation building, and which are not.

- This is called L1 Regularization since we add penality equivalent to absolute value of magnitude of coefficents.
- Due to value of penality being absolute magnitude can become 0.

$$\boxed{\text{Lasso Regression cost function} = \text{cost function} + \lambda \sum_{i=1}^{n} |\theta_n|}$$

[C] Elastic-Net Regularization

- Combination of Ridge and Lasso Regularization

$$\boxed{\text{Elastic Net Regularization cost function} = \text{Cost function} + \lambda \sum_{i=1}^{n} (\theta_n)^2 + \lambda \sum_{i=1}^{n} |\theta_n|}$$

Impact of Regularization