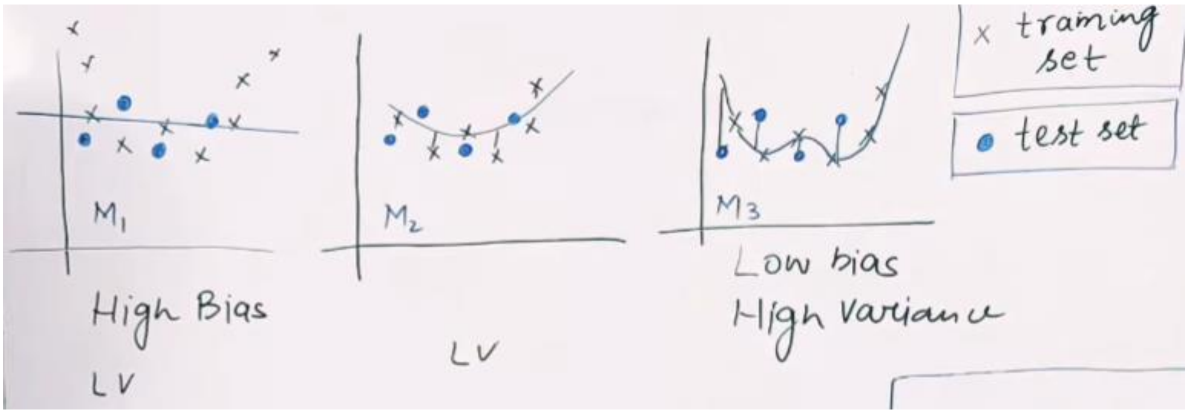
Machine Learning

Video 61:

Bias Variance Trade-off | Overfitting and Underfitting in Machine Learning

The Bias-Variance Trade-off in machine learning refers to the balance between two types of errors: bias (error due to overly simplistic models) and variance (error due to models that are too complex and overfit). A good model minimizes both bias and variance, achieving generalization without underfitting or overfitting.



Underfitting Overfitting

We will look for: Low bias, low variance

Since, in above example, only the middle one satisfies both, thus selected.

Video 62:

Ridge Regression Part 1 | Geometric Intuition and Code | Regularized Linear Models

The three common techniques to reduce overfitting are:

1. **Regularization**: Adding a penalty term (e.g., L1 or L2) to the model's loss function to constrain the complexity of the model.
2. **Cross-validation**: Using techniques like k-fold cross-validation to evaluate the model's performance on different data subsets, ensuring it generalizes well.
3. **Pruning (for decision trees)**: Reducing the size of the tree by removing branches that have little importance, preventing the model from becoming too complex.

What is Ridge Regularization?

Ridge regularization, also known as **L2 regularization**, is a technique used to prevent overfitting in linear models (like linear regression). It adds a penalty term to the loss function based on the sum of the squared values of the model's coefficients. This encourages the model to keep the coefficients small, which reduces model complexity and prevents overfitting.

Ridge regression is particularly useful when there are many correlated features, as it helps to reduce their impact by shrinking the coefficients towards zero.

Mathematically, the Ridge regression loss function looks like this:

$$\text{Loss} = \text{Sum of Squared Errors} + \lambda \times \sum_{i=1}^{n} \theta_i^2$$

Where:

- λ is the regularization parameter (controls the strength of regularization),
- θ_i are the model's coefficients.

What is overfitting?

Overfitting occurs when a machine learning model learns not only the underlying patterns in the training data but also the noise or random fluctuations. As a result, the model performs very well on the training data but poorly on unseen test data because it has become too complex and tailored to the specificities of the training set. Overfitting leads to poor generalization, meaning the model struggles to make accurate predictions on new, unseen data.

Signs of overfitting include:

- High accuracy on training data, but low accuracy on test data.
- Model being excessively complex (e.g., too many features or parameters).

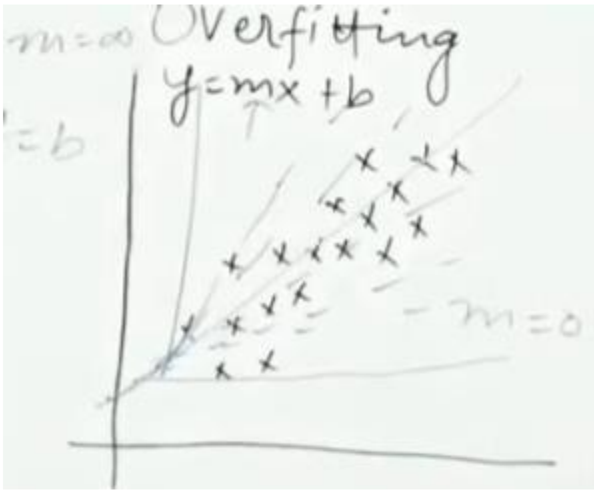
What is overfitting in terms of Ridge regularization?

In terms of **Ridge regularization**, **overfitting** happens when a linear model becomes too complex and learns noise or irrelevant patterns from the training data. Without regularization, the model may fit the training data very closely, resulting in low training error but poor generalization to new, unseen data. This overfitting is problematic because the model essentially "memorizes" the training data rather than learning the true underlying relationships.

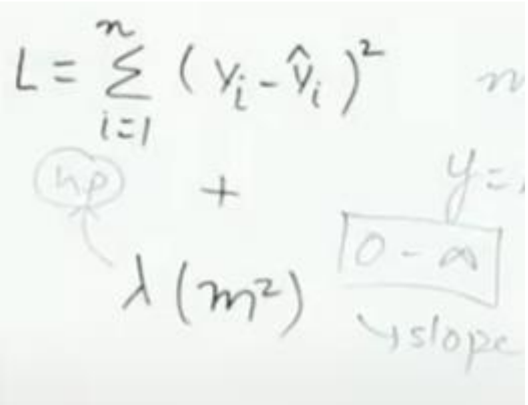
Ridge regularization addresses overfitting by adding a penalty term to the loss function. This term discourages large coefficients for the features, effectively "shrinking" them towards zero. By controlling the size of the coefficients, Ridge regularization prevents the model from becoming excessively complex and helps it generalize better to new data.

In Ridge regression, overfitting can occur when the regularization parameter (λ\lambda) is too small, allowing the model to retain large coefficients. Increasing λ\lambda helps mitigate overfitting by imposing a stronger penalty, leading to a simpler model that performs better on unseen data.

In layman terms, y = mx + b, is overfitting if m is too large.



We know we have to minimize:



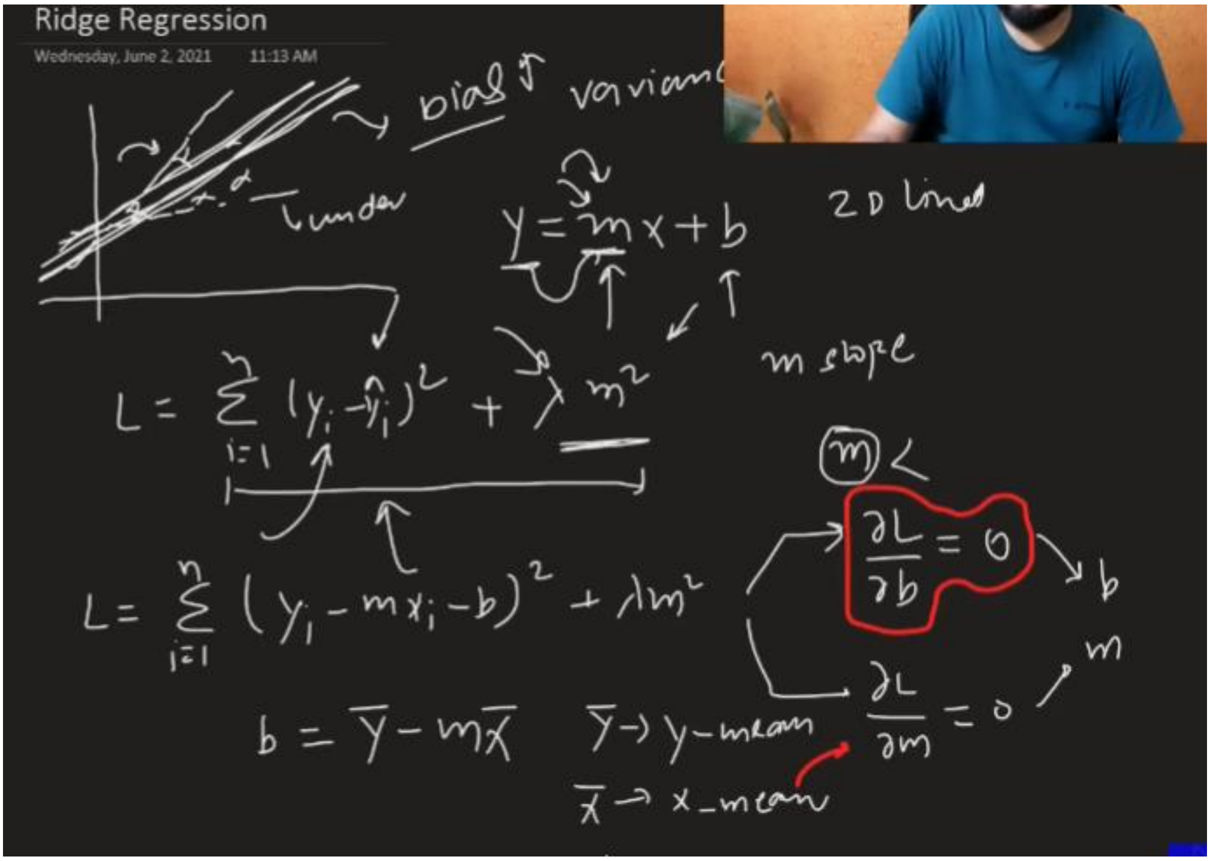
Example:

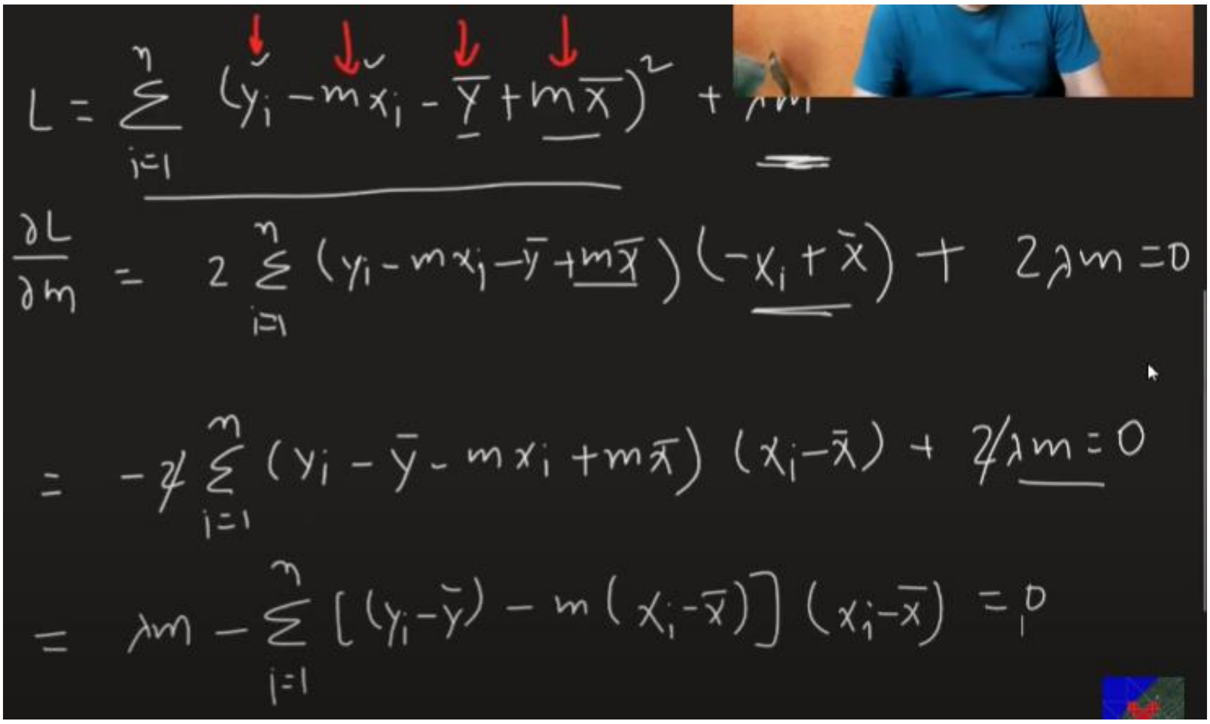
Code link:

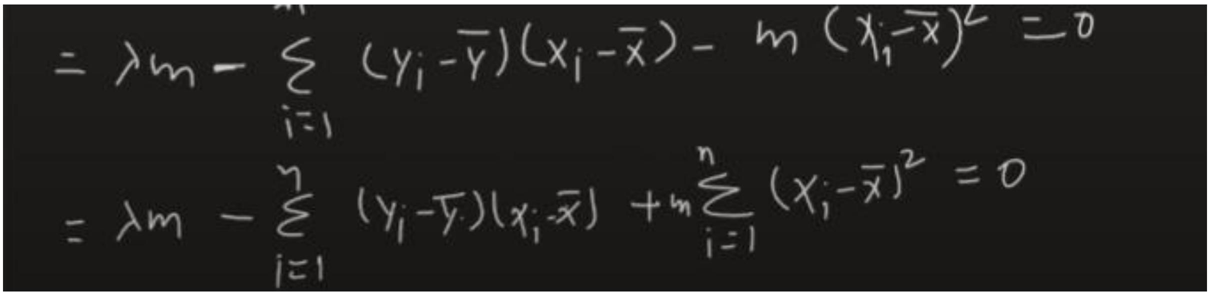
https://colab.research.google.com/drive/1fls4qbDL_jpHDSTybQlR_LcAcegnb8MM?usp=sharing

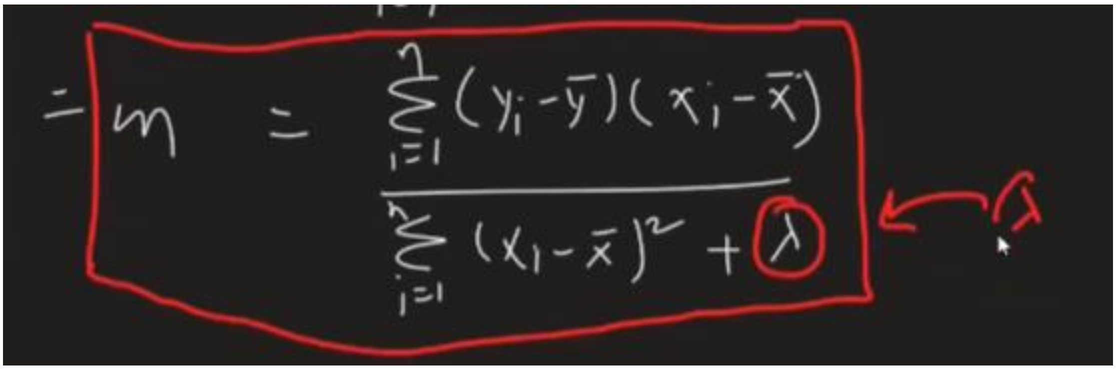
Video 63:

Ridge Regression Part 2 | Mathematical Formulation & Code from scratch | Regularized Linear Models







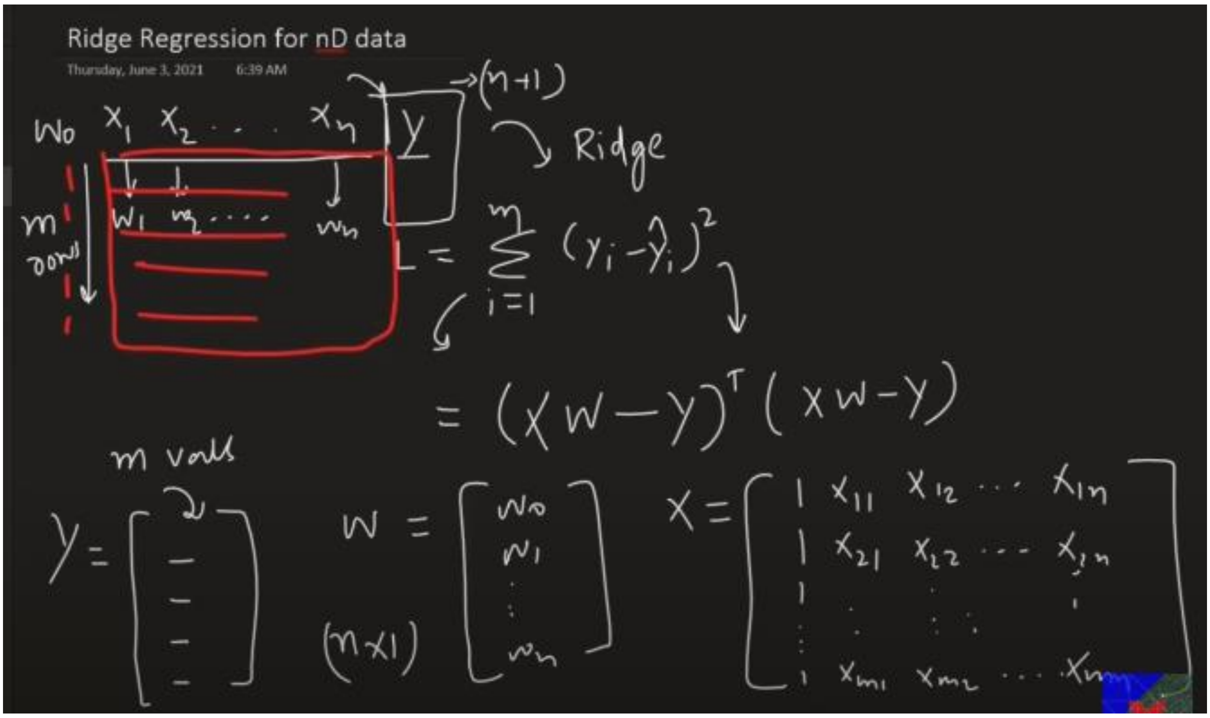


Example:

Code link:

https://colab.research.google.com/drive/1fls4qbDL_jpHDSTybQlR_LcAcegnb8MM?usp=sharing

Ridge regression for n D data:



Example:

Code link:

https://colab.research.google.com/drive/1fls4qbDL_jpHDSTybQlR_LcAcegnb8MM?usp=sharing

Video 64:

Ridge Regression Part 3 | Gradient Descent | Regularized Linear Models

Example:

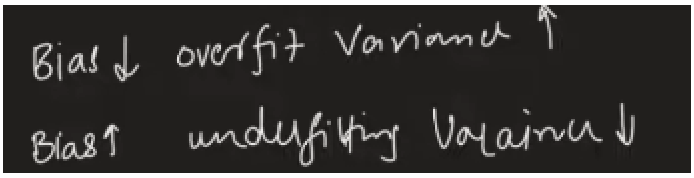
Code link:

https://colab.research.google.com/drive/1kjv9DCG1fpSrKGY5E9gxbJcHUeeGWGSa?usp=sharing

Video 65:

5 Key Points - Ridge Regression | Part 4 | Regularized Linear Models

1. How the coefficients get affected? They move close towards 0 (but not equal to 0)
2. Higher values are impacted more by increasing the value of alpha
3. Bias variance trade off



4. Effect of regularization on loss function – shifts
5. Why called ridge?