

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Regularization

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Regularization seeks to solve model issues by:

1. Minimizing model complexity
2. Penalizing the loss function
3. add more bias to reduce model variance
4. optimal penalty hyperparameter

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Three main types of Regularization:

- L1 Regularization (LASSO Regression)
- L2 Regularization (Ridge Regression)
- Combining L1 and L2 (Elastic Net)

Led by : Shreyas Shukla

L1 regularization adds a penalty which is equal to the **absolute value of the magnitude of coefficients.**

- Limits the size of the coefficients.
- Can yield sparse models where some coefficients can become zero.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

L1 regularization adds a penalty which is equal to the **absolute value of the magnitude of coefficients.**

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j|$$

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Remember:

- All coefficients are shrunk by the same factor.
- Does not necessarily eliminate coefficients.

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2$$

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

L2 regularization adds a penalty equal to the **square** of the magnitude of coefficients.

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2$$

Led by : Shreyas Shukla

Elastic net combines L1 and L2 with the addition of an alpha parameter deciding the ratio between them:

$$\frac{\sum_{i=1}^n (y_i - x_i^J \hat{\beta})^2}{2n} + \lambda \left(\frac{1 - \alpha}{2} \sum_{j=1}^m \hat{\beta}_j^2 + \alpha \sum_{j=1}^m |\hat{\beta}_j| \right)$$

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

These regularization methods do have a cost:

1. Introduce an additional hyperparameter that needs to be tuned.
2. A multiplier to the penalty to decide the “strength” of the penalty.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Before we dive into coding, let's discuss a few more relevant topics:

- Feature Scaling
- Cross Validation

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Feature Scaling

Led by : Shreyas Shukla

- Some ML models that rely on distance metrics (e.g. KNN) **require** scaling to perform well.
- Benefits:
 1. FS improves the convergence of steepest descent algorithms, which do not possess the property of scale invariance.
 2. If features are on different scales, certain weights may update faster than others since the feature values $\mathbf{x_j}$ play a role in the weight updates.

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- Also important in comparing measurements with different units.
-
- Allows direct comparison of model coefficients.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Some rules:

- Must always scale new unseen data before feeding to model.
- Effects direct interpretability of feature coefficients

Easier to compare coefficients to one another, harder to relate back to original unscaled feature.

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- Feature scaling benefits:
 - Can lead to great increases in performance.
 - Absolutely necessary for some models.
 - Virtually no “real” downside to scaling features.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- Two main ways to scale features:
 - Standardization: Rescales data to have a mean (μ) of 0 and standard deviation (σ) of 1
 - Normalization : Rescales all data values to be between 0-1

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Standardization:

Rescales data to have a mean (μ) of 0 and standard deviation (σ) of 1 (unit variance).

$$X_{changed} = \frac{X - \mu}{\sigma}$$

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

No confusion Please

Standardization also referred to as “Z-score normalization”.

$$X_{changed} = \frac{X - \mu}{\sigma}$$

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Normalization: Scales all data values to be between 0 and 1.

iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

$$X_{changed} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Led by : Snreyas Snukia

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Let's discuss the fit and transform calls in scaling.

iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- A `.fit()` method call simply calculates the necessary statistics (Xmin,Xmax,mean, standard deviation).
- A `.transform()` call actually scales data and returns the new scaled version of data.
- Previously saw a similar process for polynomial feature conversion.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Very important:

- We only **fit** to training data.
- Calculating statistical information should only come from training data.
- Don't want to assume prior knowledge of the test set!
- No **Data Leakage** please !!

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Using the full data set would cause **data leakage**:

- Calculating statistics from full data leads to some information of the test set leaking into the training process upon transform() conversion.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- Feature scaling process:
 - Perform train test split
 - Fit **ONLY** to training feature data
 - Transform training feature data
 - Transform test feature data

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- Do we need to scale the label?
- not necessary nor advised.
- Can negatively impact stochastic gradient descent.
- stochastic gradient descent is advanced topic. Help Yourself !!

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- Let's move to cross-validation!

iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

Led by : Shreyas Shukla