

Evaluation Regression Model



Regression Model	Pros	Cons
Linear Regression	Works on any size of dataset, gives informations about relevance of features	The Linear Regression Assumptions
Polynomial Regression	Works on any size of dataset, works very well on non linear problems	Need to choose the right polynomial degree for a good bias/variance tradeoff
SVR	Easily adaptable, works very well on non linear problems, not biased by outliers	Compulsory to apply feature scaling, not well known, more difficult to understand
Decision Tree Regression	Interpretability, no need for feature scaling, works on both linear / nonlinear problems	Poor results on too small datasets, overfitting can easily occur
Random Forest Regression	Powerful and accurate, good performance on many problems, including non linear	No interpretability, overfitting can easily occur, need to choose the number of trees

Regularization

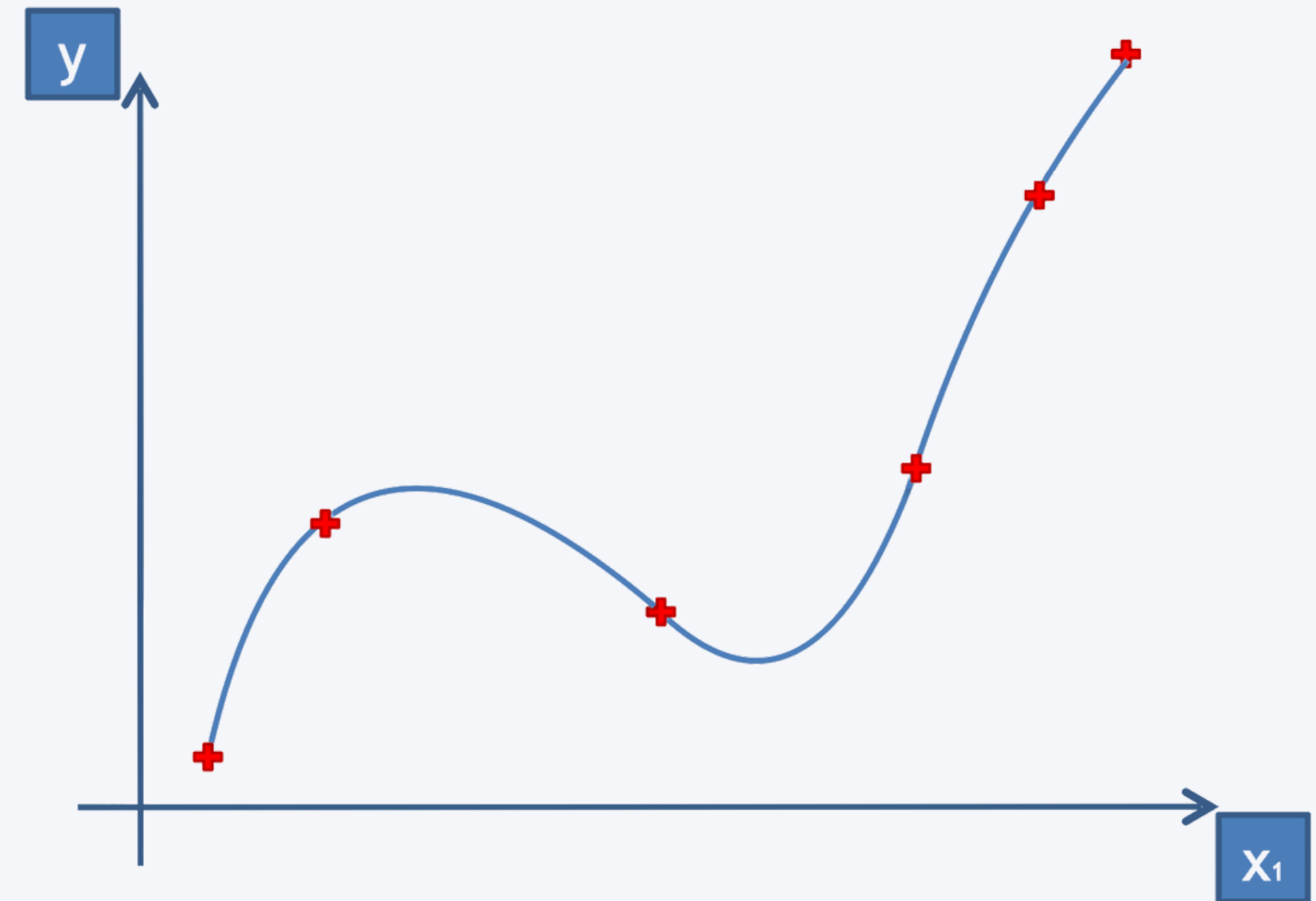
Methods

The Problem of Overfitting

- A curve perfectly fitting all data points
- The model captures noise instead of the real pattern.

Key Points:

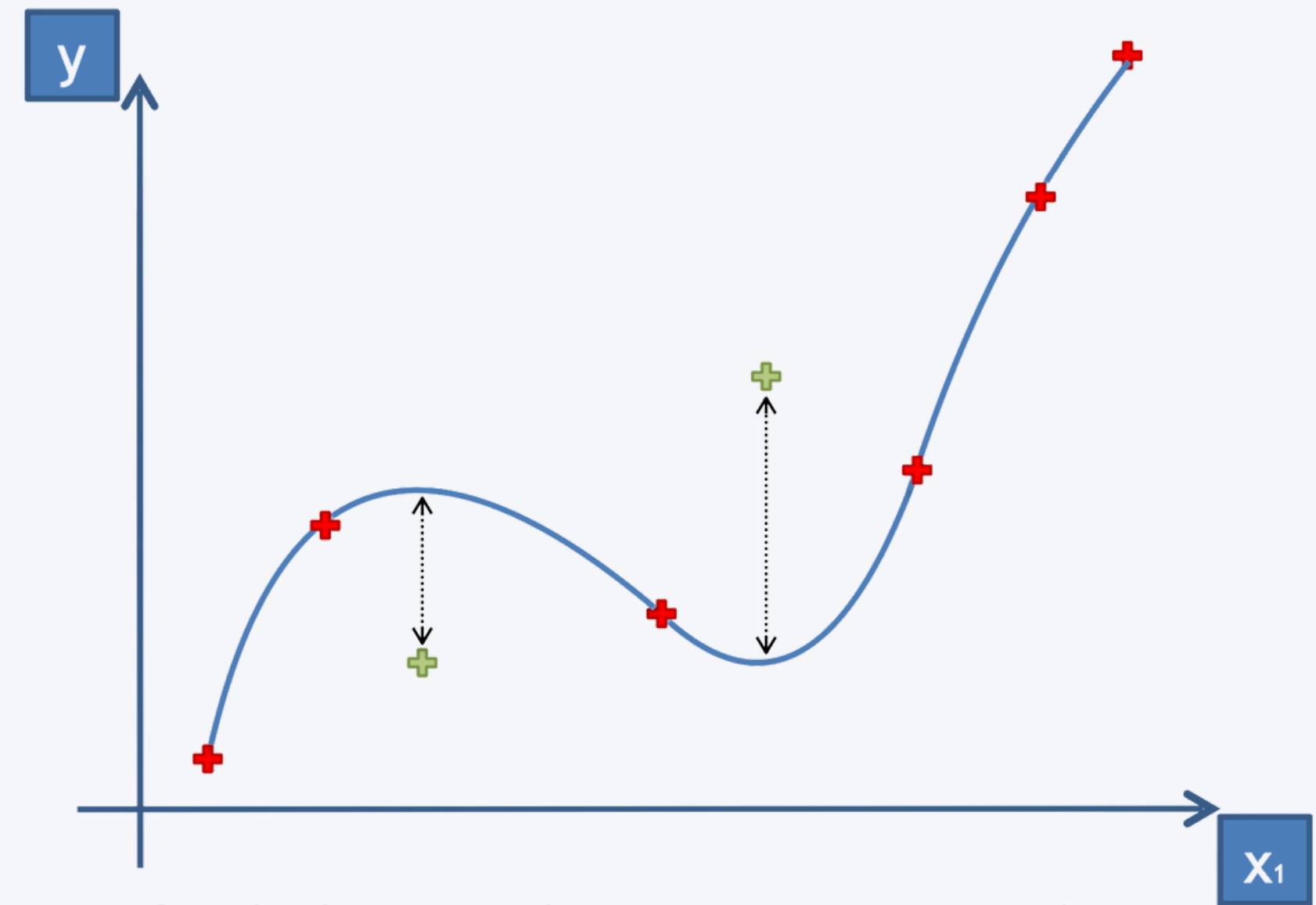
- **Overfitting:** The model is too complex, learning noise instead of trends.
- **Effects:** High training accuracy but poor test performance.
- **Indicators:** High variance, unstable predictions, poor generalization.



This fitting curve fits the data perfectly well

The Problem of Overfitting

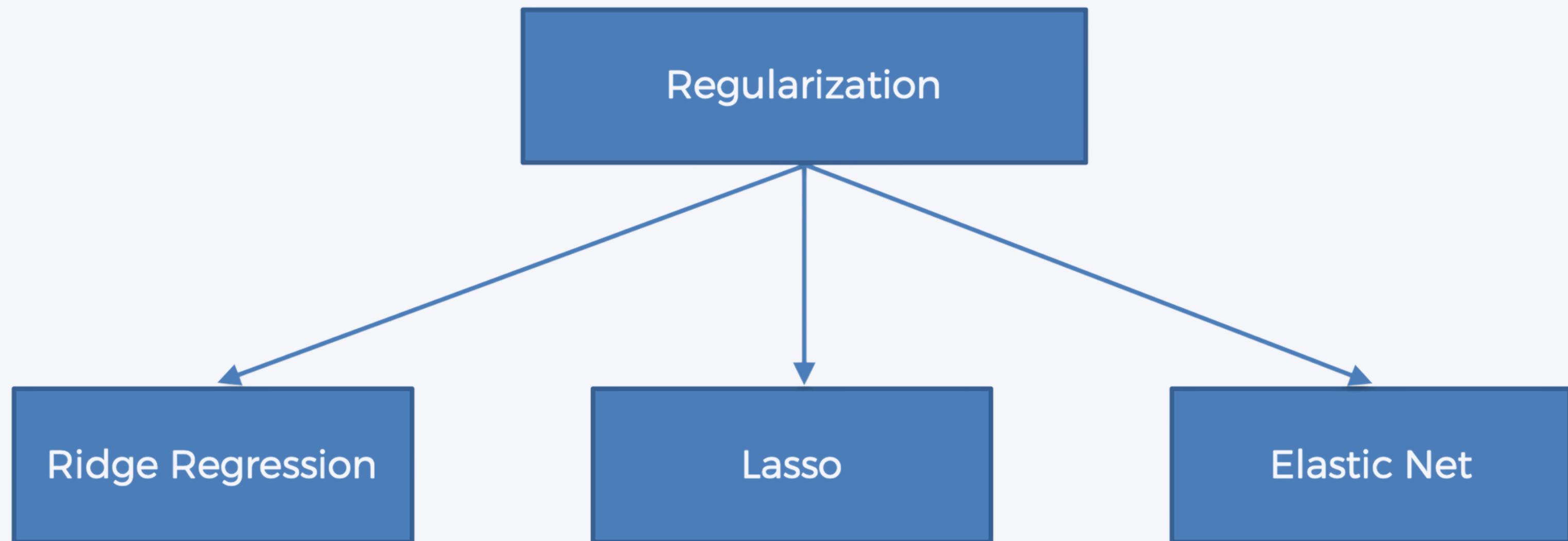
Small changes in input lead to big prediction shifts.



But if we look at new observations, we can get large errors

Examples of Regularization

Regularization is a technique in regression that adds a penalty to the model to prevent overfitting and improve generalization.



Without Regularization

This is the **standard Mean Squared Error (MSE)** minimization in linear regression.

- The model aims to minimize the sum of squared differences between actual values y_i and predicted values.

Key Points:

- **No Penalty on Coefficients:** The model freely adjusts parameters b_0, b_1, \dots, b_m .
- **Risk of Overfitting:** If there are too many features, the model may fit noise instead of patterns.

$$\text{Minimize } \sum_{i=1}^n (y^i - (b_0 + b_1 x_1^i + \dots + b_m x_m^i))^2$$

Ridge Regression

- Ridge regression adds a **penalty (λ * sum of squared coefficients)** to the standard loss function.
- This penalty shrinks coefficients but does not force them to zero.

$$\text{Minimize } \sum_{i=1}^n (y^i - (b_0 + b_1 x_1^i + \dots + b_m x_m^i))^2 + \lambda(b_1^2 + \dots + b_m^2)$$

Key Points:

- Controls Overfitting by reducing coefficient magnitudes.
- Smooths model predictions and improves generalization.
- λ (lambda) controls regularization strength – higher λ shrinks coefficients more.

Lasso Regression

- Lasso adds a penalty on the **absolute values of coefficients**
 - ($|b_1| + |b_2| + \dots + |b_m|$).
- Encourages sparsity by forcing some coefficients to exactly zero.

$$\text{Minimize } \sum_{i=1}^n (y^i - (b_0 + b_1 x_1^i + \dots + b_m x_m^i))^2 + \lambda(|b_1| + \dots + |b_m|)$$

Key Points:

- **Feature Selection:** Automatically removes less important features.
- **Sparse Models:** Useful when only a few features matter.
- **λ (lambda)** controls regularization strength – higher λ forces more coefficients to zero.

Elastic Net

Elastic Net combines **Lasso (L1)** and **Ridge (L2)** penalties.

- The penalty includes both absolute values and squared values of coefficients.

$$\text{Minimize } \sum_{i=1}^n (y^i - (b_0 + b_1 x_1^i + \dots + b_m x_m^i))^2 + \lambda_1 (|b_1| + \dots + |b_m|) + \lambda_2 (b_1^2 + \dots + b_m^2)$$

Key Points:

- **Best of Both Worlds:** Shrinks coefficients like **Ridge** and performs feature selection like **Lasso**.
- Useful when features are **correlated**: Lasso alone may arbitrarily drop one feature, but Elastic Net keeps important ones.
- **Two Hyperparameters (λ_1, λ_2):** Control the balance between L1 and L2 regularization.

Regularization Methods

As a Result,

- Regularization **prevents** overfitting by simplifying the model.
- **Better generalization:** The model performs well on both training and new data.
- **Less variance:** Predictions are more stable and robust.

