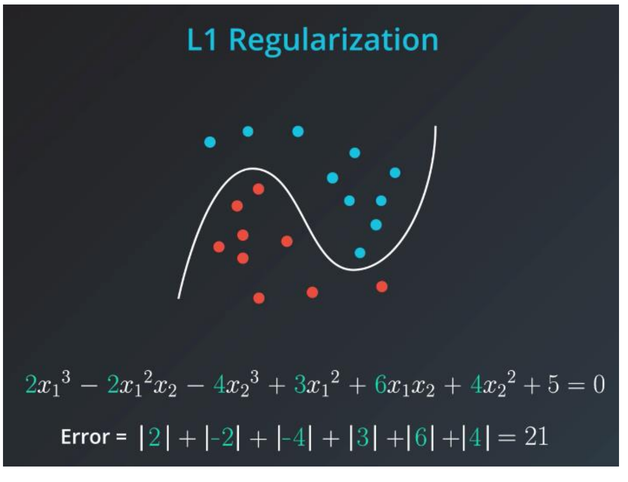
Regularization

A useful technique to improve our models and make sure that they don't overfit.

Regularization is a way to take the complexity of a model into account when computing the error.

L1-Regularization

We take all coefficients, sum up their absolute values and add them to the error.



L2-Regularization

Same idea, but instead of adding values we add the squares of the coefficients.

require more complexity, while others, e.g. a video recommender leaves more room for complexity.

it's common to introduce a parameter that allows us to influence the "strength" of punishment by multiplication with the computed regularization factor.(lambda)

L1- vs. L2- Regularization

L1 Regularization	L2 Regularization
Computationally Inefficient (unless data is sparse)	Computationally Efficient
Sparse Outputs	Non-Sparse Outputs
Feature Selection	No Feature Selection

Ridge Regression and Lasso Regression are both regularization techniques used in machine learning to prevent overfitting and improve the performance of a model.

Both Ridge and Lasso aim to add a penalty term to the standard linear regression objective function to control the complexity of the model.

Ridge Regression (L2 Regularization): Adds the sum of squared magnitudes of coefficients as a penalty term.

Lasso Regression (L1 Regularization): Adds the sum of absolute magnitudes of coefficients as a penalty term.

Ridge Regression: Tends to shrink the coefficients towards zero, but they rarely become exactly zero.

Lasso Regression: Tends to shrink some coefficients all the way to zero, effectively performing feature selection by eliminating certain features.

Feature Selection:

Ridge Regression: Does not perform feature selection; all features are retained but with reduced magnitudes.

Lasso Regression: Can lead to sparse models by effectively eliminating some features.