Regularization Part 3:

Elastic-Net Regression

We ended the StatQuest on **Lasso Regression** by saying that it works best when your model contains a lot of useless variables…

the sum of the squared residuals

$$+$$

$$\lambda \times |\text{variable}_1| + \ldots + |\text{variable}_x|$$

We also said that **Ridge Regression** works best when most of the variables in your model are useful.

the sum of the squared residuals

$$+$$

$$\lambda \times \text{variable}_1^2 + \ldots + \text{variable}_x^2$$

Great! When we know a lot about all of the parameters in our model, it's easy to choose if we want to use **Lasso Regression**…

the sum of the squared residuals

$$+$$

$$\lambda \times |\text{variable}_1| + \ldots + |\text{variable}_x|$$

…or **Ridge Regression**….

the sum of the squared residuals

$$+$$

$$\lambda \times \text{variable}_1^2 + \ldots + \text{variable}_x^2$$

…but what do we do when we have a model that includes tons more variables?

Size = y-intercept + slope_1 × **Weight** + diet difference × **High Fat Diet** + slope_2 × **Age** + slope_3 × **Size of Father** + …. + **tons more variables**…

So how do you choose if you should use **Lasso** or **Ridge Regression**?

Size = y-intercept + slope$_1$ × **Weight** + diet difference × **High Fat Diet** + slope$_2$ × **Age** + slope$_3$ × **Size of Father** + …. + **tons more variables**…

The good news is that you don't have to choose, instead, you use **Elastic-Net Regression!!!**

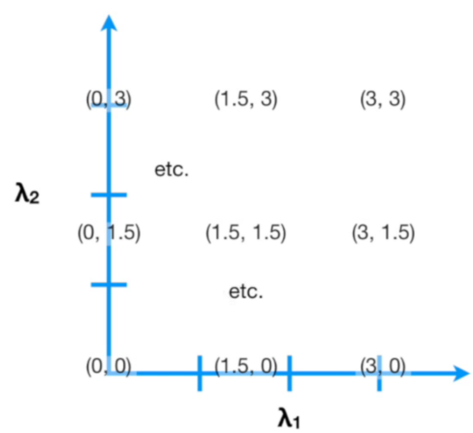
Altogether, **Elastic Net Regression** combines the strengths of **Lasso** and **Ridge Regression**.

the sum of the squared residuals

+

$\boldsymbol{\lambda_1} \times |\textbf{variable}_1| + \ldots + |\textbf{variable}_X| \quad \mathbf{+} \quad \boldsymbol{\lambda_2} \times \textbf{variable}_1^2 + \ldots + \textbf{variable}_X^2$

We use **Cross Validation** on different combinations of $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ to find the best values.



The hybrid **Elastic-Net Regression** is especially good at dealing with situations when there are correlations between parameters.

the sum of the squared residuals

$$+$$

$$\lambda_1 \times |\text{variable}_1| + \ldots + |\text{variable}_X| \;\; + \;\; \lambda_2 \times \text{variable}_1^2 + \ldots + \text{variable}_X^2$$

This is because on it's own, **Lasso Regression** tends to pick just one of the correlated terms and eliminates the others…

the sum of the squared residuals

$$+$$

$$\lambda_1 \times |\text{variable}_1| + \ldots + |\text{variable}_X|$$

…whereas **Ridge Regression** tends to shrink all of the parameters for the correlated variables together.

the sum of the squared residuals

$$+$$

$$\lambda_2 \times \text{variable}_1^2 + \ldots + \text{variable}_X^2$$

By combining **Lasso** and **Ridge Regression**, **Elastic-Net Regression** groups and shrinks the parameters associated with the correlated variables and leaves them in equation or removes them all at once.

the sum of the squared residuals

$$+$$

$$\lambda_1 \times |\text{variable}_1| + \ldots + |\text{variable}_X| \;\; + \;\; \lambda_2 \times \text{variable}_1^2 + \ldots + \text{variable}_X^2$$