Regularization for Simplicity: L₂ Regularization

Consider the following **generalization curve**, which shows the loss for both the training set and validation set against the number of training iterations.

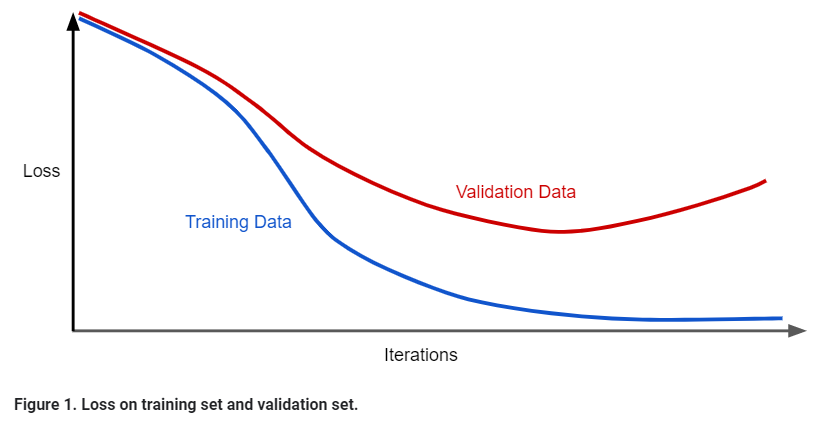
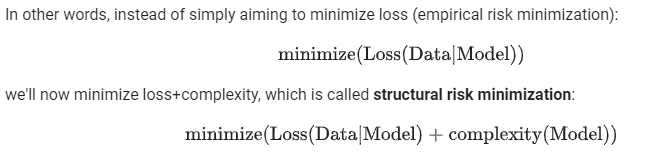


Figure 1 shows a model in which training loss gradually decreases, but validation loss eventually goes up. In other words, this generalization curve shows that the model is [overfitting](https://developers.google.com/machine-learning/crash-course/generalization/peril-of-overfitting) to the data in the training set. Channeling our inner [Ockham](https://developers.google.com/machine-learning/crash-course/generalization/peril-of-overfitting#ockham), perhaps we could prevent overfitting by penalizing complex models, a principle called **regularization**.



Our training optimization algorithm is now a function of two terms: the **loss term**, which measures how well the model fits the data, and the **regularization term**, which measures model complexity.

Machine Learning Crash Course focuses on two common (and somewhat related) ways to think of model complexity:

* Model complexity as a function of the *weights* of all the features in the model.
* Model complexity as a function of the *total number of features* with nonzero weights. (A [later module](https://developers.google.com/machine-learning/crash-course/regularization-for-sparsity/l1-regularization) covers this approach.)

If model complexity is a function of weights, a feature weight with a high absolute value is more complex than a feature weight with a low absolute value.

