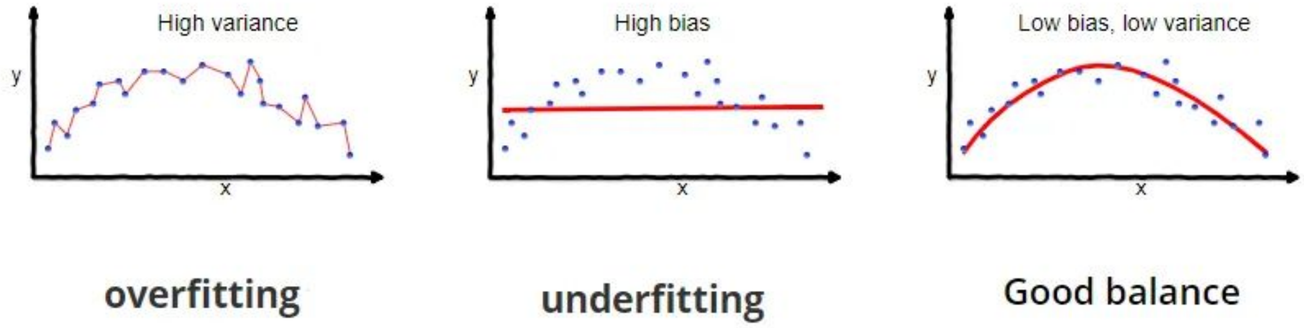
Model Complexity and Regularization

Bias vs Variance Tradeoff

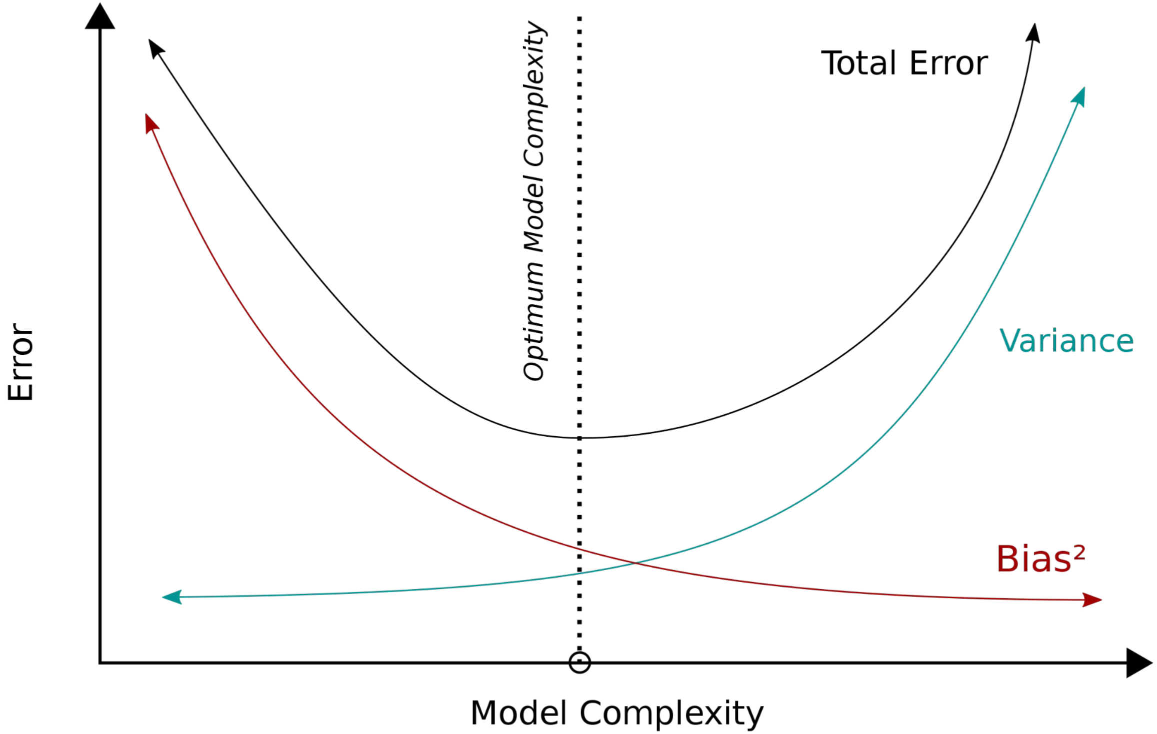


Bias: What is the inherent error that you obtain from your classifier even with infinite training data? This is due to your classifier being "biased" to a particular kind of solution (e.g. linear classifier). In other words, bias is inherent to your model.

Variance: Captures how much your classifier changes if you train on a different training set. How "over-specialized" is your classifier to a particular training set (overfitting)? If we have the best possible model for our training data, how far off are we from the average classifier?

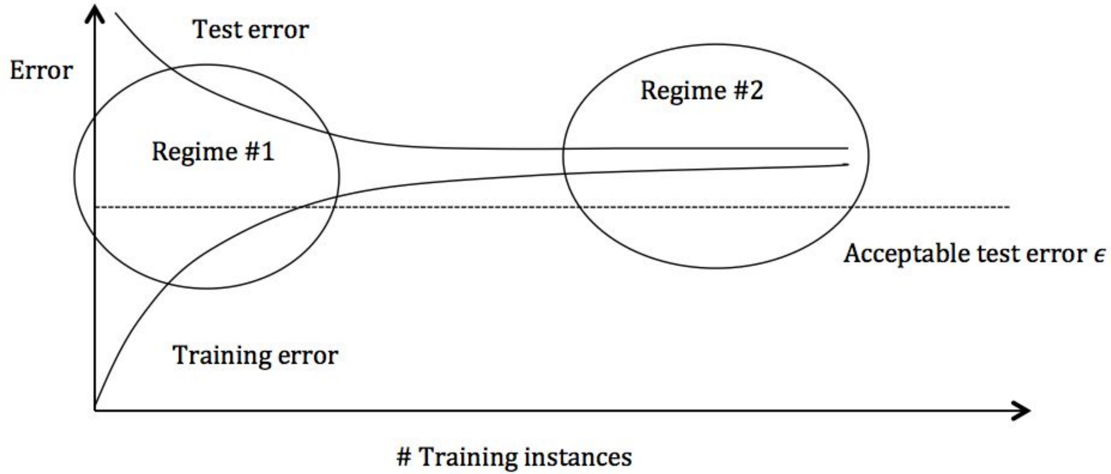
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229

Bias vs Variance Tradeoff



https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff

Bias vs Variance Tradeoff



Regime 1 (High Variance)
Symptoms:
1. Training error is much lower than test error
2. Training error is lower than ϵ
3. Test error is above ϵ

Remedies:
- Add more training data
- Reduce model complexity (e.g. regularization) -- complex models are prone to high variance
- Bagging

Regime 2 (High Bias)
Symptoms:
1. Training error is higher than ϵ

Remedies:
- Use more complex model (e.g. kernelize, use non-linear models)
- Add features
- Boosting

https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote12.html

Multivariate Linear Regression

Linear Model:

$$\hat{y} = w[0] \times x[0] + w[1] \times x[1] + \ldots + w[n] \times x[n]$$

Cost Function:

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} \left(y_i - \sum_{j=0}^{p} w_j \times x_{ij} \right)^2$$

Multivariate Linear Regression + Regularization

Ridge Regression (L2):

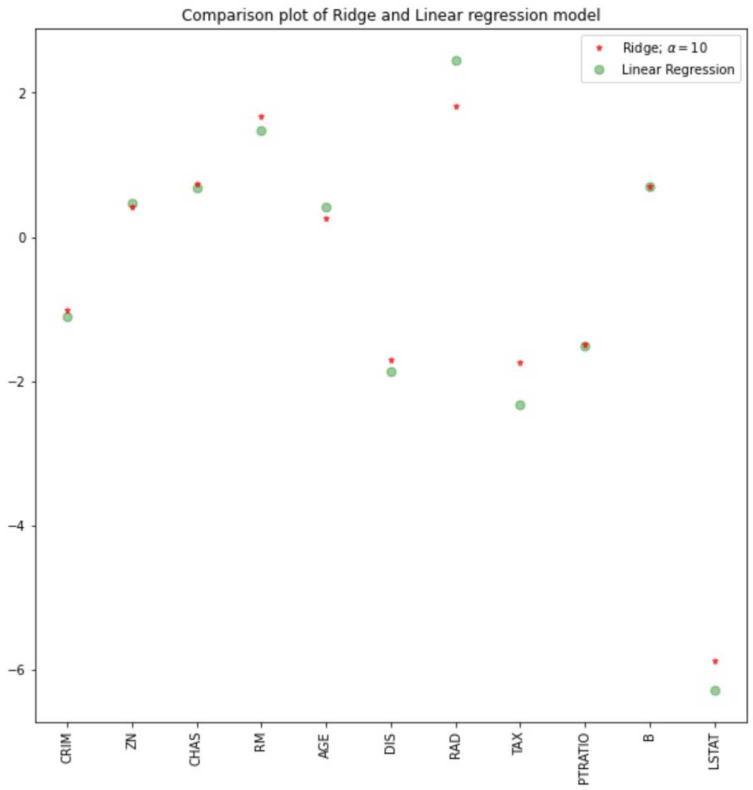
$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 \boxed{+ \lambda \sum_{j=0}^{p} w_j^2}$$

Lasso Regression (L1):

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 \boxed{+ \lambda \sum_{j=0}^{p} |w_j|}$$

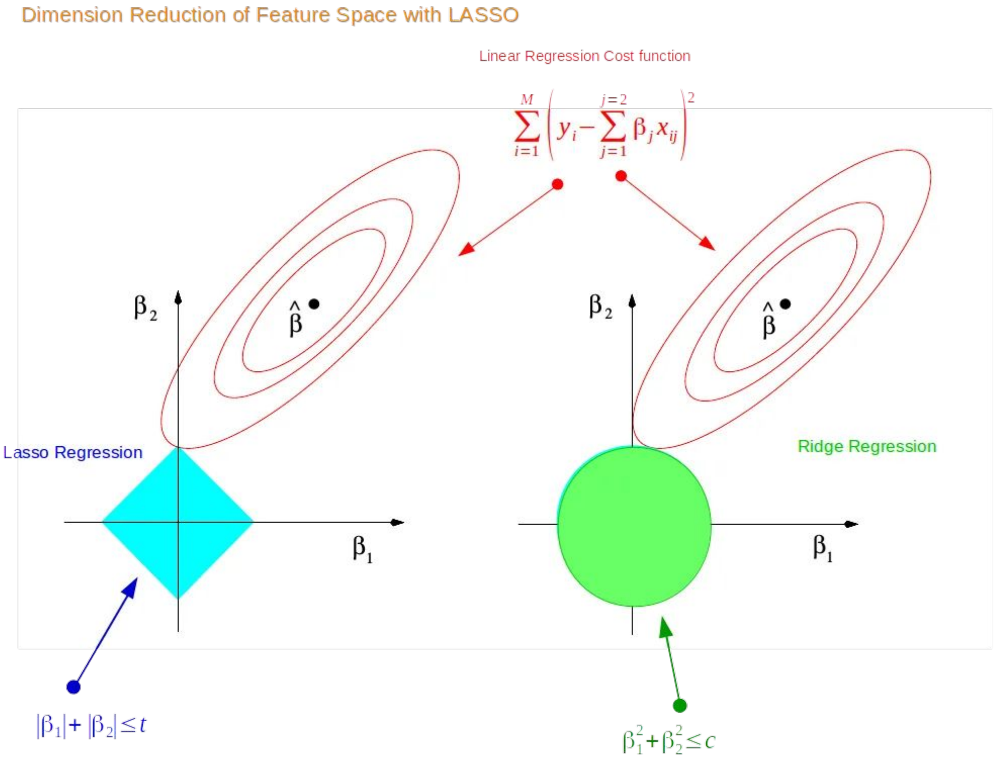
Goal: Minimize coefficient weights (lower model complexity) to reduce overfitting and increase generalizability

Multivariate Linear Regression + Regularization



https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression

Multivariate Linear Regression + Regularization



https://towardsdatascience.com/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b