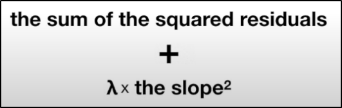
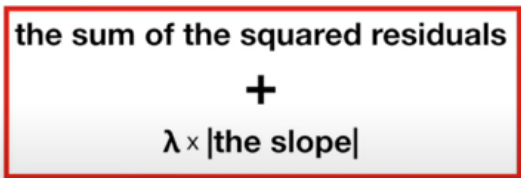
Regularization

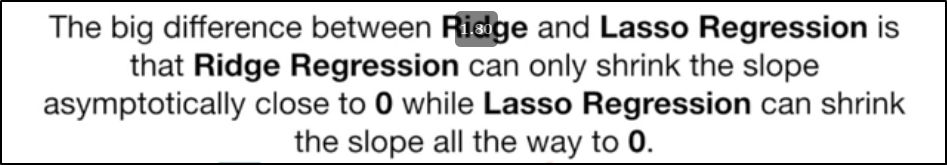
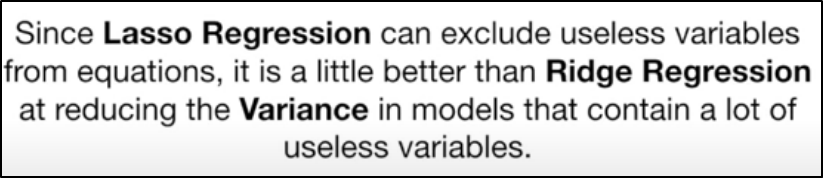
## [Ridge (L2) Regression](https://youtu.be/Q81RR3yKn30?list=PLblh5JKOoLUICTaGLRoHQDuF_7q2GfuJF)

1. **Regularization = Optimized Desensitization**
   1. find a line that doesn’t fit train so well **(introduce bias) to decrease variance** (diff in fit between train and test)
2. The L2 lines seeks to minimize 🡪
3. Parts of the Equation
   1. **Lambda** = Strength of Penalty
      1. If Lambda = 0, equation becomes least squares (standard regression)
      2. As Lambda gets larger and larger outcome variable gets less and less sensitive to the predictor variable
   2. **[Squared] Penalty** = slope^2 + var.1.residuals^2 + var.2.residuals^2 …
      1. All var residuals
      2. Not all vars are desensitized/shrunk equally
      3. Does NOT include y-intercept
4. [L2 also works with discrete vars](https://youtu.be/Q81RR3yKn30?list=PLblh5JKOoLUICTaGLRoHQDuF_7q2GfuJF&t=630)
5. [L2 can be applied to Logistic Regression](https://youtu.be/Q81RR3yKn30?list=PLblh5JKOoLUICTaGLRoHQDuF_7q2GfuJF&t=804)
   1. Note: Sum of likelihoods is optimized, not squared residuals
6. [AMAZING: Ridge Regression can solve for parameters (lambda, penalty) when there isn’t enough data](https://youtu.be/Q81RR3yKn30?list=PLblh5JKOoLUICTaGLRoHQDuF_7q2GfuJF&t=1112)
7. [SUMMARY](https://youtu.be/Q81RR3yKn30?list=PLblh5JKOoLUICTaGLRoHQDuF_7q2GfuJF&t=1155)

## [Lasso (L1) Regression](https://youtu.be/NGf0voTMlcs?list=PLblh5JKOoLUICTaGLRoHQDuF_7q2GfuJF)

Minimizes **[Absolute] Penalty**



1. Differences from L2
2. 
3. 
4. 