Applied Machine Learning

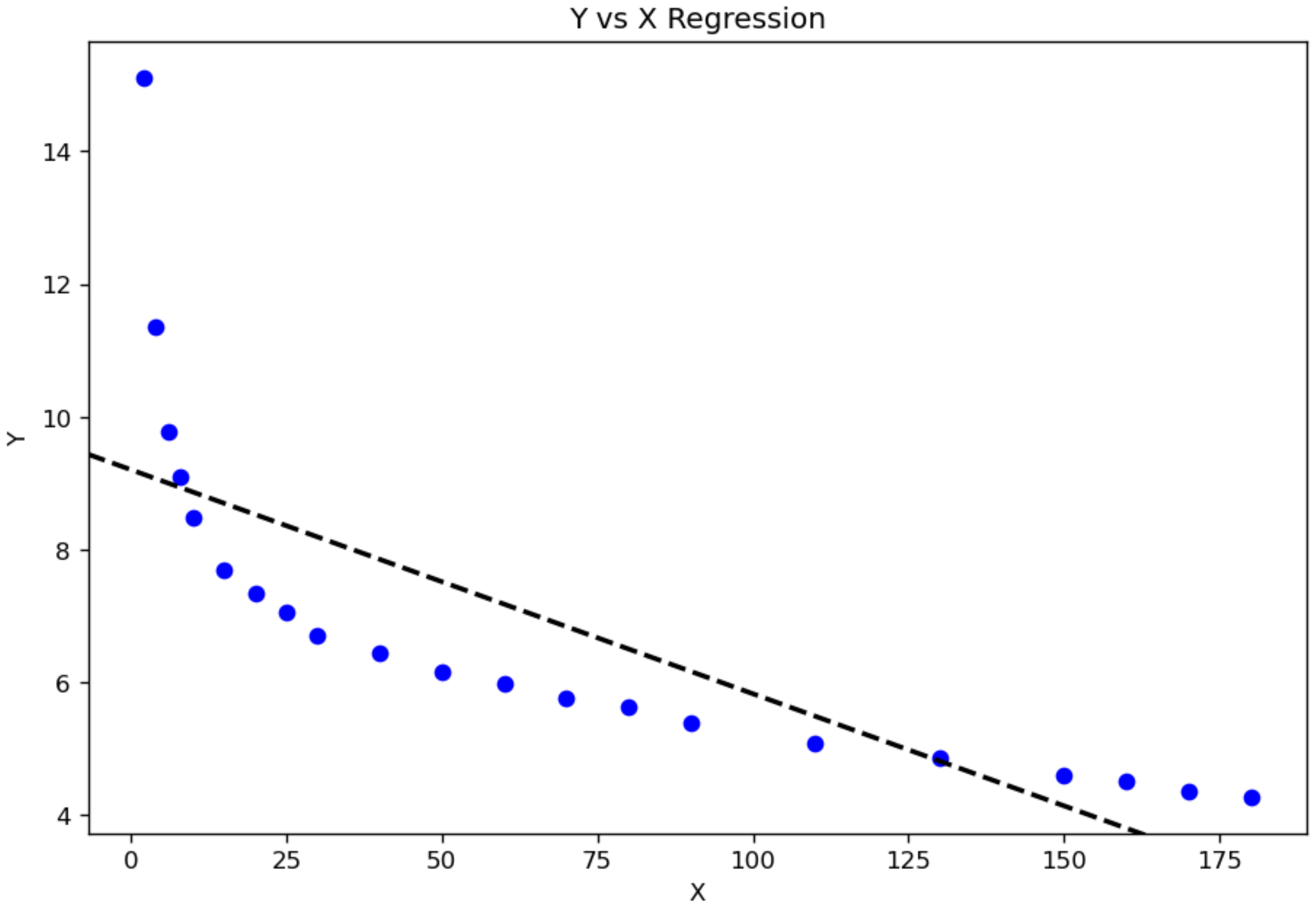
Linear Regression - Transformation of Variables

Linear Regression - Transformations

- Box-Cox transformation
- Polynomials of the same explanatory variable
- Regularization

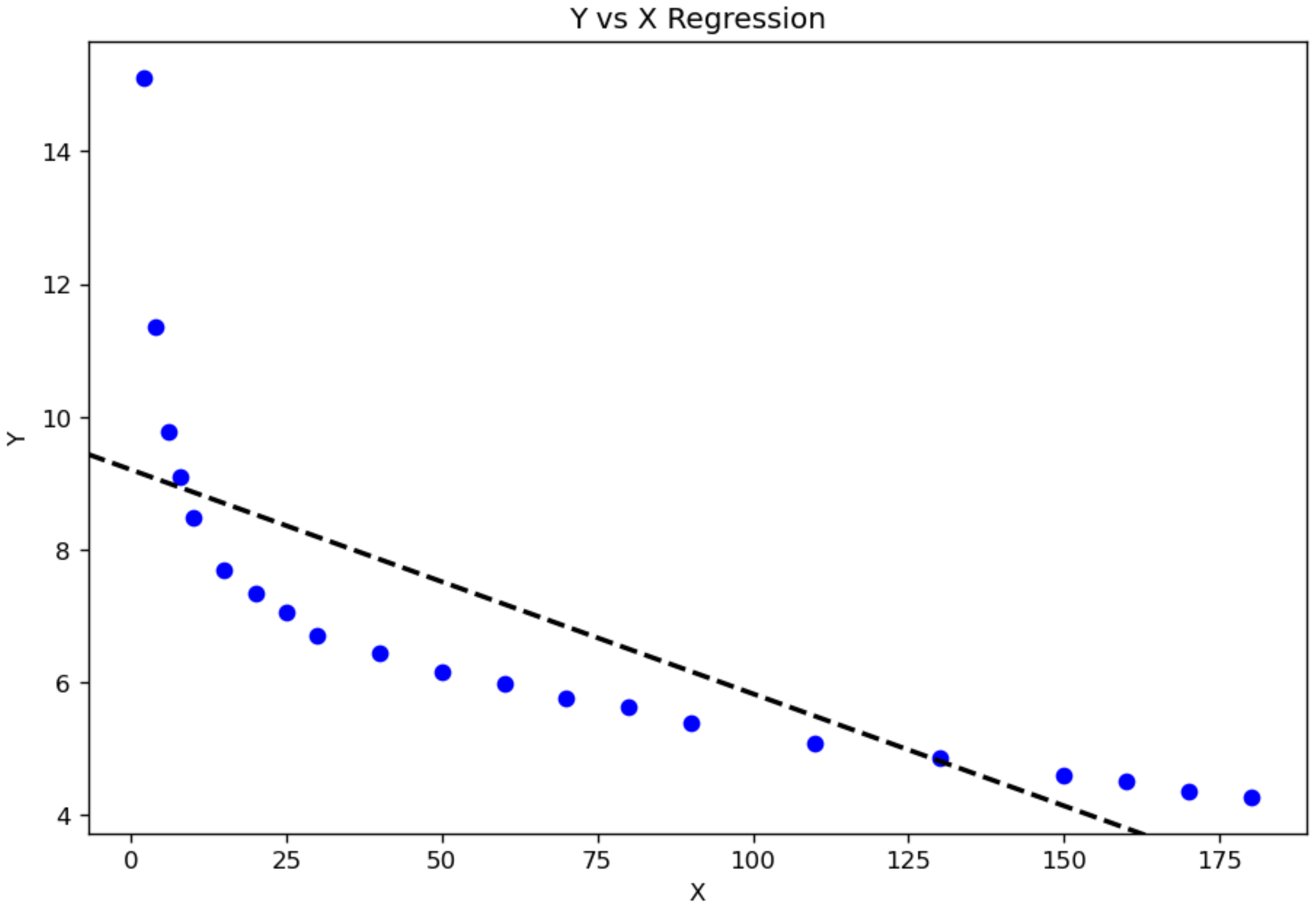
Linear Regression

- Low R^2
    - Linear function may not be best to explain dataset
- A transformation may help
    - box-cox transformation of explanatory or dependent variables
    - polynomials of explanatory variables



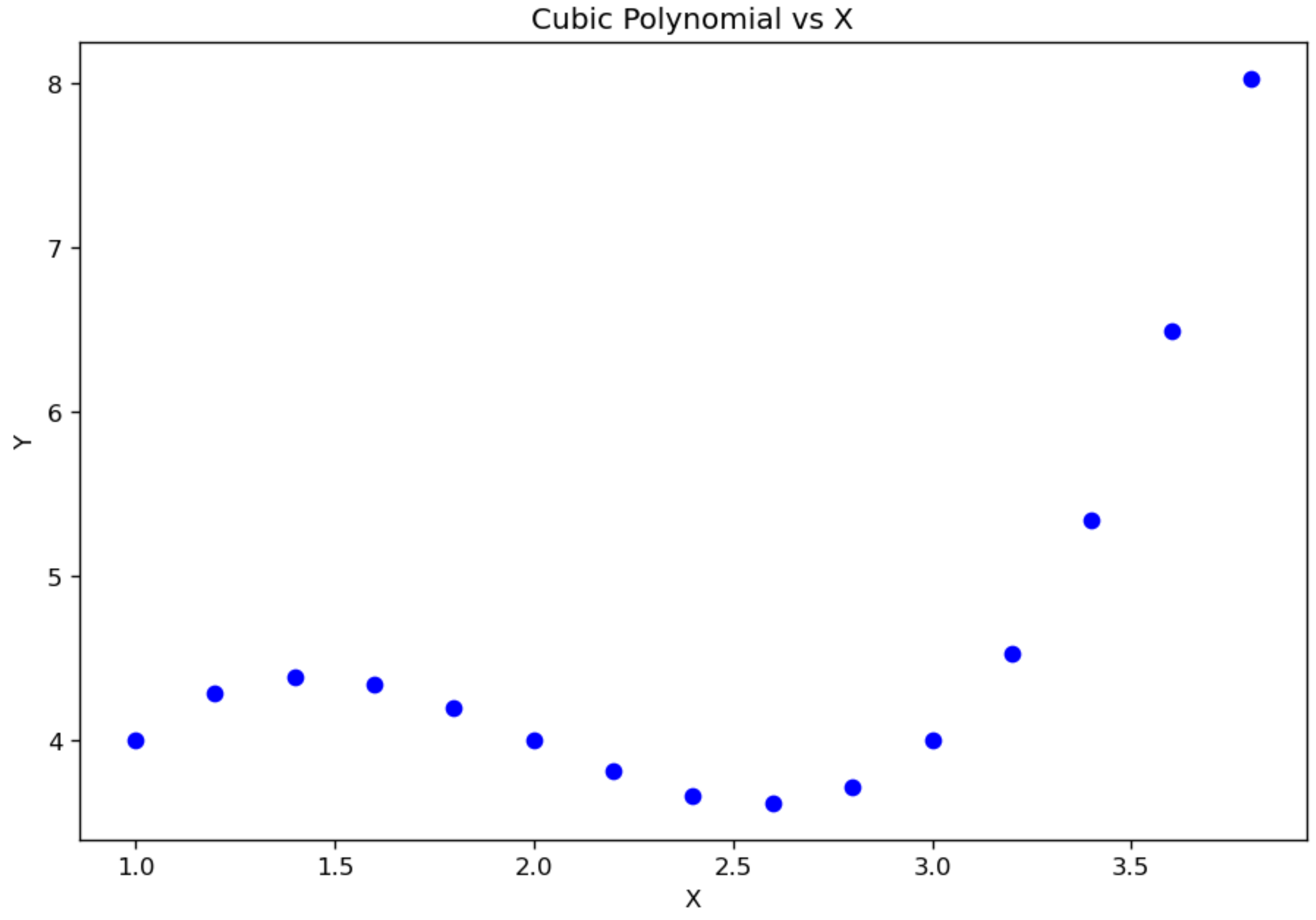
Box-Cox Transformation

- $$y_i^{(bc)} = \begin{cases} \dfrac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln y_i & \text{if } \lambda = 0 \end{cases}$$
  - defined for $y_i \geq 0$
  - statistical libraries to search for λ
- Inverse transformation:
- $$y_i = \begin{cases} |\lambda y_i^{(bc)} + 1|^{\frac{1}{\lambda}} & \text{if } \lambda \neq 0 \\ e^{y_i^{(bc)}} & \text{if } \lambda = 0 \end{cases}$$



Polynomial of Explanatory Variable

- Lineal model of 2 explanatory variables
  - $y_i = \beta_2 x_2^{(i)} + \beta_1 x_1^{(i)} + \beta_0 + \xi^{(i)}$
  - $X_i^\top = [x_2^{(i)}, x_1^{(i)}, 1]$
- Polynomial model of order 2 with 1 explanatory variable
  - $y_i = \beta_2 (x^{(i)})^2 + \beta_1 (x^{(i)})^1 + \beta_0 + \xi^{(i)}$
  - $X_i^\top = [(x^{(i)})^2, (x^{(i)})^1, 1]$
- Apply regression as before
- Hard to determine appropriate order



Regularization

- Regression solves
  - $X^\top X\hat{\beta} = X^\top \mathbf{y}$
- Correlation among explanatory variables
  - small eigenvectors of $X^\top X$
    - $X^\top X\hat{\beta} \approx X^\top X(\hat{\beta} + \mathbf{w})$ for large $\mathbf{w}$
    - large values in $\hat{\beta}$ yield to large errors in prediction

Regularization

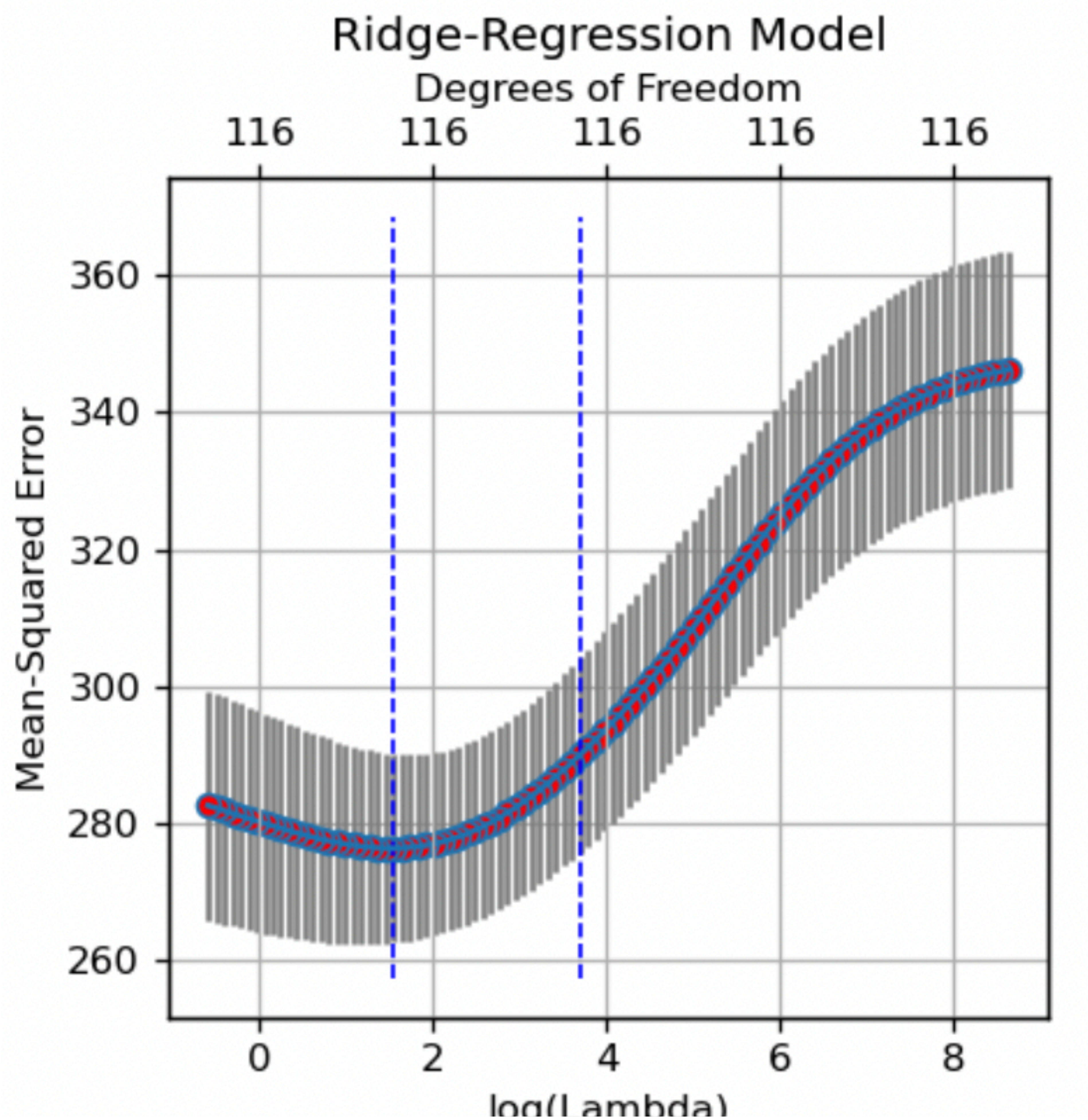
- Regression solves
  - $X^\top X\hat{\beta} = X^\top \mathbf{y}$
- Large values $\hat{\beta}$ yield to large errors in prediction
- Add regularization term to penalize larve values of β and minimize:
  - $\frac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^\top\beta$
  - $\lambda \geq 0$

Regularization

- Add regularization term and minimize
  - $\frac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^\top\beta$
  - $\lambda \geq 0$ penalizes large values of β
- $\beta^\top\beta$ is the L_2 norm of β: Ridge Regression
- Differentiating with respect to β and equating to 0. Regression:
  - $\left[\frac{1}{N}\mathbf{X}^\top\mathbf{X} + \lambda I\right]\beta = \frac{1}{N}\mathbf{X}^\top\mathbf{y}$

Regularization - Finding λ

- Cross-validation
  - consider choices of λ at different scales, e.g., $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$
- for each λ_i,
  - iteratively build new random Fold from Training Set
    - fit Cross-Validation Train Set using λ_i
    - compute MSE for current Fold on Validation set
  - for each λ_i record average error and σ over all Folds
- λ with smallest error - largest λ within one σ



Regularization - Putting it all together

Fit dependent variable from explanatory variables

- outliers
  - standardized residuals, leverage, Cook’s distance
- not a random normal variable or predictions suggest non-linearities
  - box-cox
  - polynomial representation
- regularization
  - cross-validation

Linear Regression - Transformations

- Box-Cox transformation
- Polynomials of the same explanatory variable
- Regularization

Applied Machine Learning

Linear Regression - Transformation of Variables