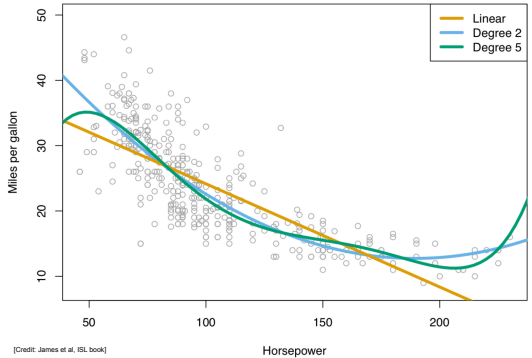
Section 3I. Nonlinear Effects

Statistics for Data Science

Victor M. Preciado, PhD MIT EECS
Dept of Electrical & Systems Engineering
University of Pennsylvania
preciado@seas.upenn.edu

Nonlinear effects

- Consider a collection of 392 vehicles. For each vehicle, we know the horsepower (input) and measure the miles per gallon (MPG), i.e., $\mathcal{D} = \{\text{MPG}_i, \text{HP}_i\}_{i=1}^{392}$
  - The relationship between MPG and HP is closer to polynomial than linear



[Credit: James et al, ISL book]

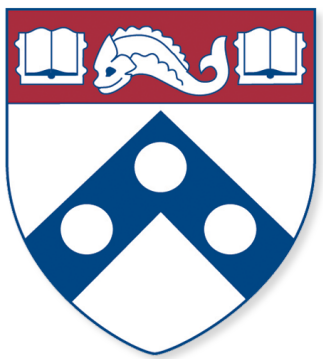
Nonlinear effects: Polynomial fit

- To predict MPG from HP, we can use a polynomial degree k:

$$\text{MPG} = \beta_0 + \beta_1 \times \text{HP} + \beta_2 \times \text{HP}^2 + \ldots + \beta_k \times \text{HP}^k + \varepsilon$$

  - We find the values of the coefficients using a simple *linear* fit with k input variables. In particular, we define the variables $\text{HP}_r = \text{HP}^r$ and perform a linear fitting using the output variable $y_i = \text{MPG}_i$ and $\mathbf{x}_i = [\text{HP}_1, \text{HP}_2, \ldots, \text{HP}_k]^\mathsf{T}$
  - The data suggest that a good fit occurs when $k = 2$ (quadratic fit)

[Credit: James et al, ISL book]	Coefficient	Std. Error	t-statistic
`Intercept`	56.9001	1.8004	31.6
`horsepower`	-0.4662	0.0311	-15.0
`horsepower`^{2}	0.0012	0.0001	10.1



Penn
Engineering