PSY9511: Seminar 2

The basics of regression and classification

Esten H. Leonardsen

26.02.2024

UNIVERSITETET
I OSLO

Outline

Today's lecture:

1. Recap of last lecture
2. Proposed solution for Assignment 1
3. Basics of regression and classification
4. Presentation of Assignment 2

The basics of regression and classification

UNIVERSITETET
I OSLO

Outline

Todays topics

- Different types of outputs y
  - Regression, classification, and other variants
- Linear regression: Restricting the scope of $\hat{f}(X)$
  - Live coding 🥳
- k-Nearest Neighbours: An intuitive solution to classification
- Logistic regression: Extending linear regression to classification
  - Live coding 🥳
- Generative models for classification

Plan for future lectures

- How do we evaluate how good our models are? (Lecture 4)
- More complex solutions to regression and classification problems (Lecture 3, 5 and onwards)

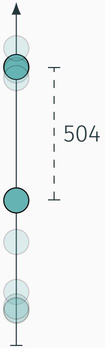
Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet

Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet

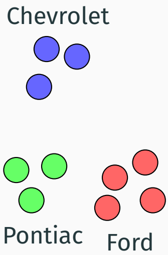
Regression vs. classification



Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet

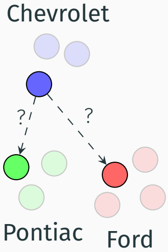
Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet



Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet



Regression vs. classification

Mean squared error (MSE):

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Accuracy:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(y_i, \hat{y}_i),$$

$$\mathbb{1}(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

Regression vs. classification

Regression:

- Predicting reaction time on a cognitive task based on sleep scores
- Predicting the age of an individual based on a brain scan
- Predicting anxiety scores based on questionnaire data

Classification:

- Predicting whether an individual is depressed based on cell phone usage data
- Predicting if a patient has dementia based on a brain scan
- Predicting whether a patient is happy based on their facial expression

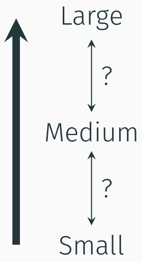
Regression vs. classification

Large

Medium

Small

Regression vs. classification

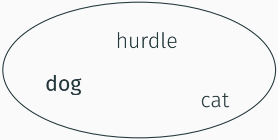


Regression vs. classification

The quick brown fox jumps over the lazy ______

Regression vs. classification

The quick brown fox jumps over the lazy ______



Classification

Regression vs. classification

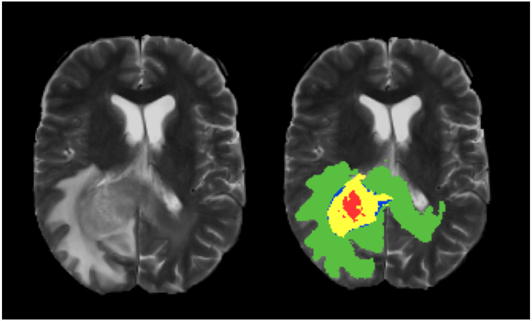
"Students taking a machine learning class" → →

Regression vs. classification

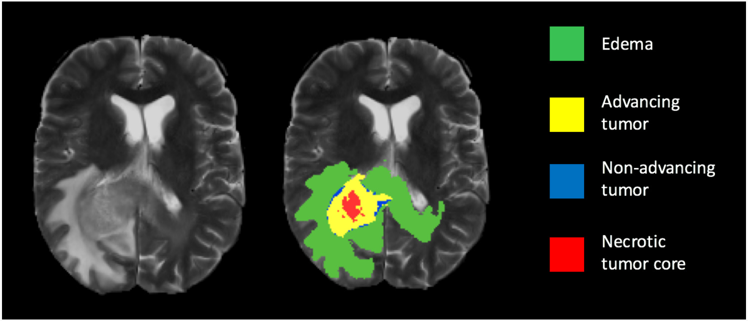
"Students taking a machine learning class"

(128, 192, 64)

Regression vs. classification



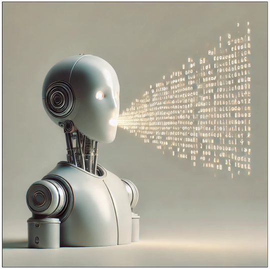
Regression vs. classification



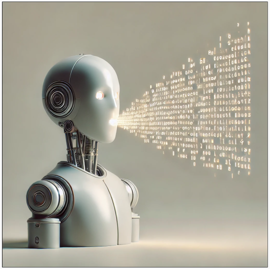
Regression vs. classification

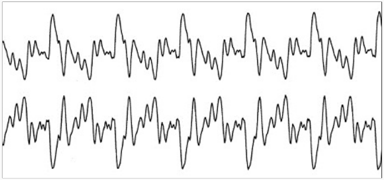
Regression vs. classification

Regression vs. classification



Regression vs. classification





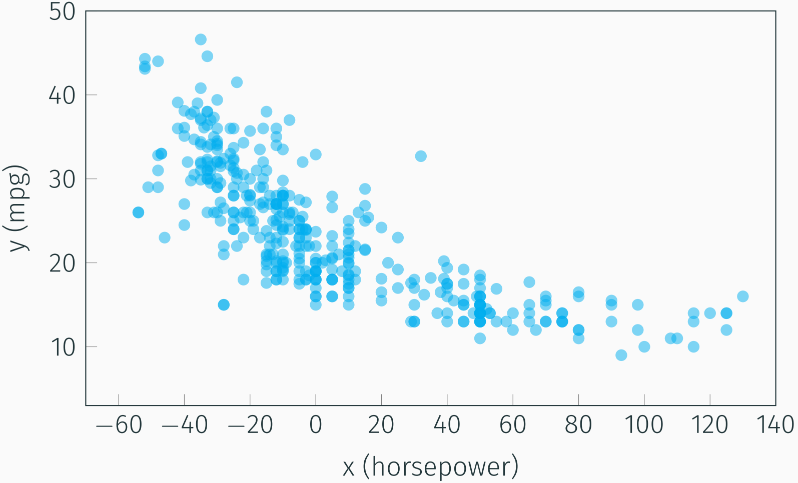
Regression vs. classification

Different types of outputs y require us to use different mathematical formulations of the problem we want to solve.

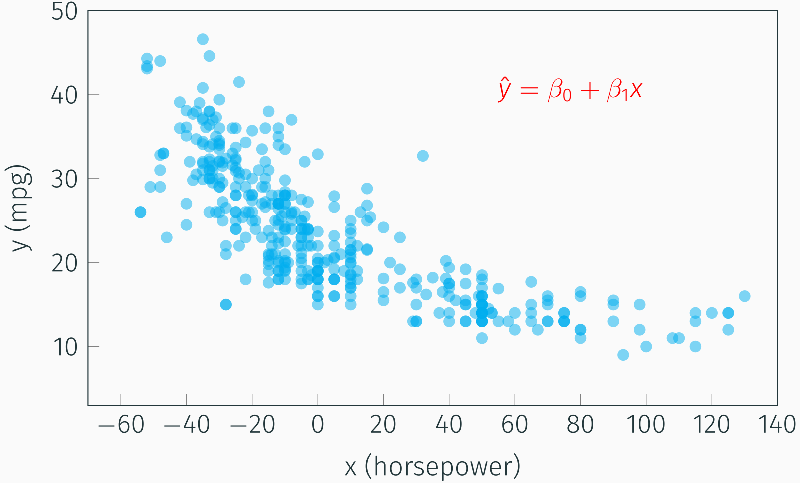
- Problems with quantitative outputs are solved via regression, often by minimizing the mean squared error
- Problems with qualitative outputs are solved by classification, often to maximize accuracy
- Ordinal regression falls between the two, with qualitative classes that have some kind of order
- A variety of other types of problems can be seen as special cases of these two

Linear regression (via ordinary least squares)

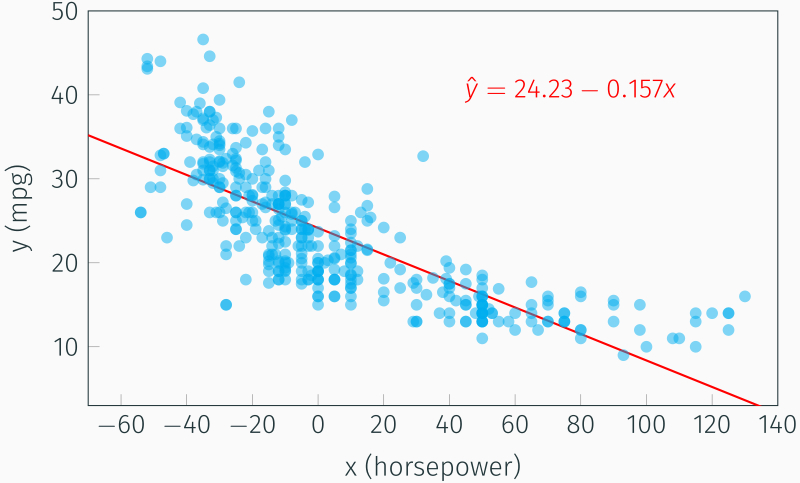
Linear regression (via ordinary least squares)



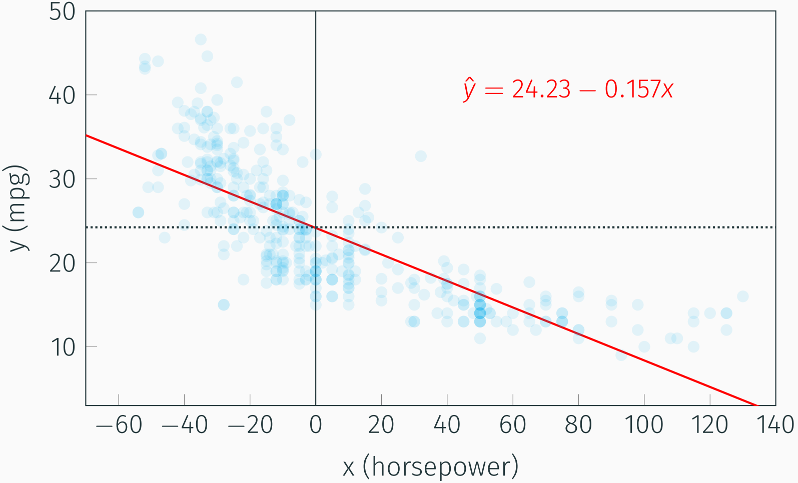
Linear regression (via ordinary least squares)



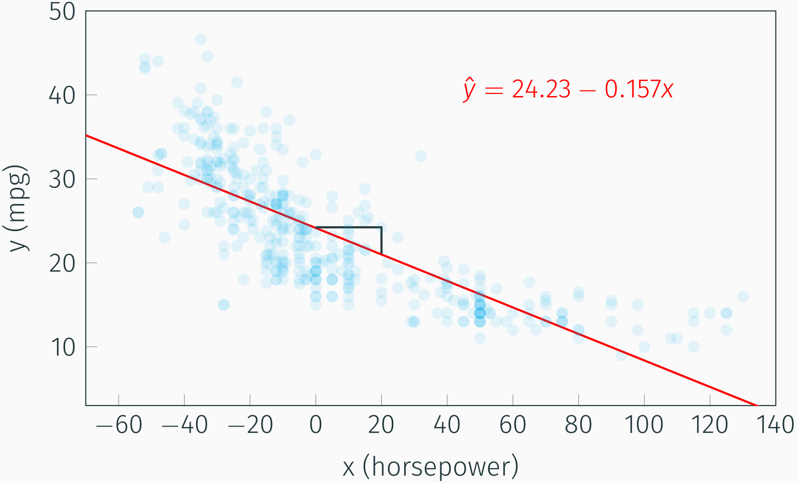
Linear regression (via ordinary least squares)



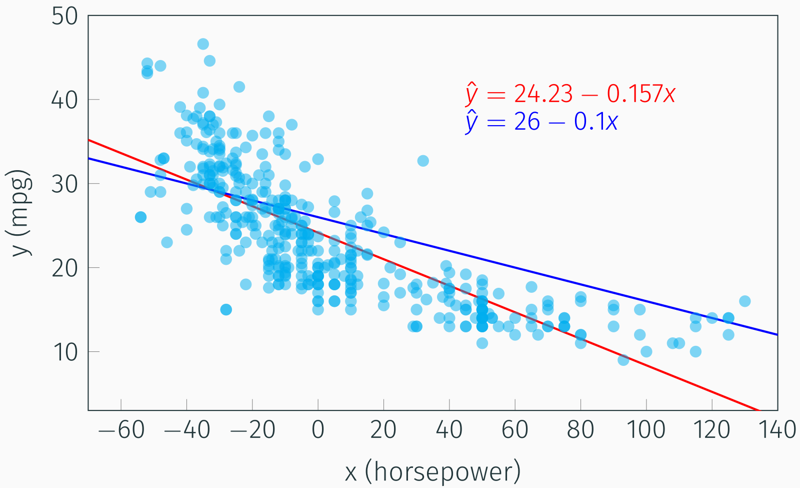
Linear regression (via ordinary least squares)



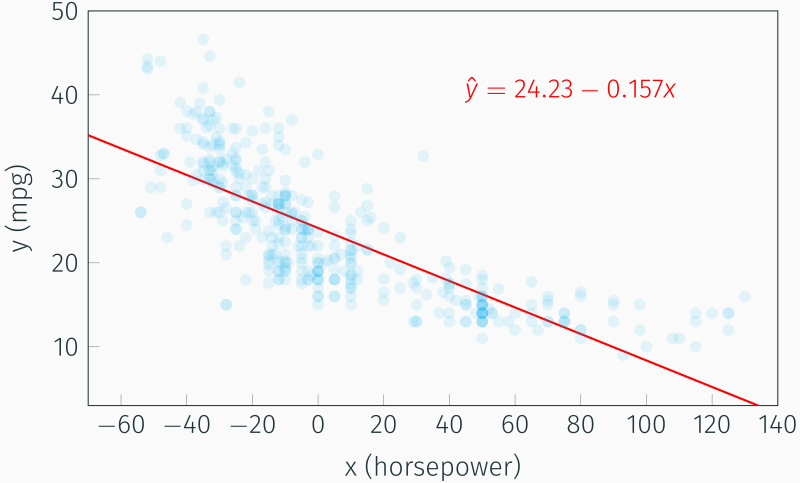
Linear regression (via ordinary least squares)



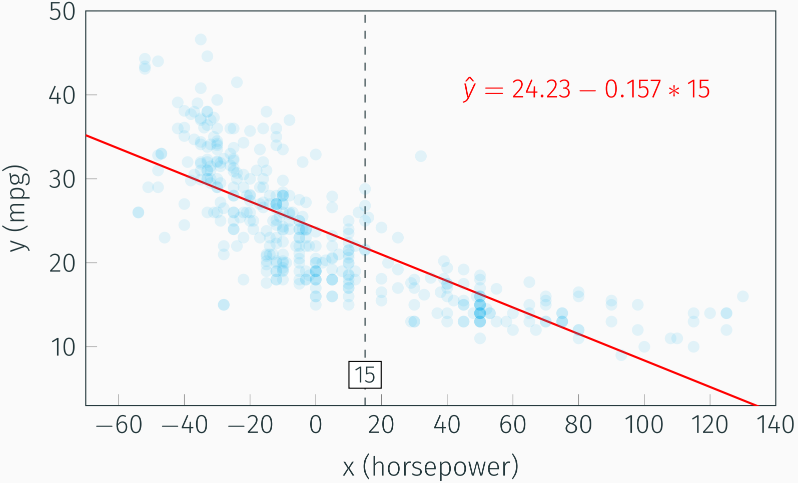
Linear regression (via ordinary least squares)



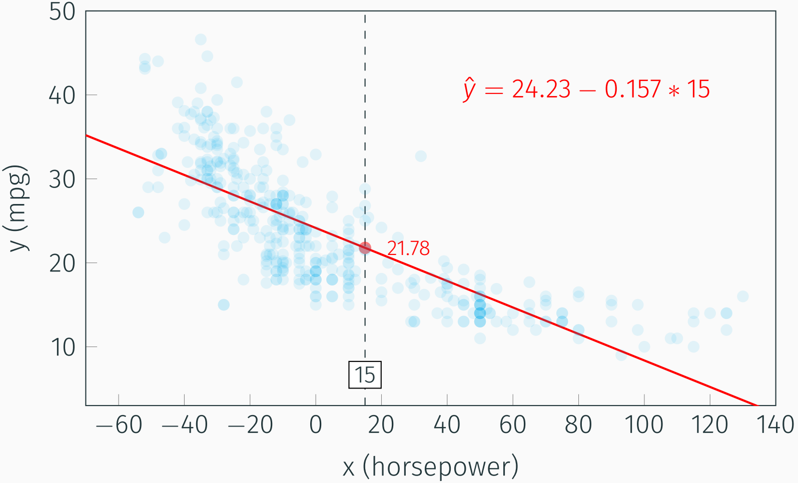
Linear regression (via ordinary least squares)



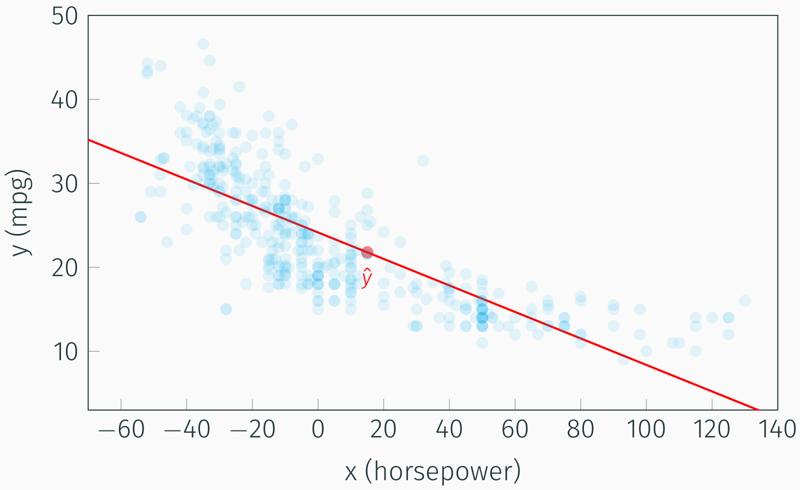
Linear regression (via ordinary least squares)



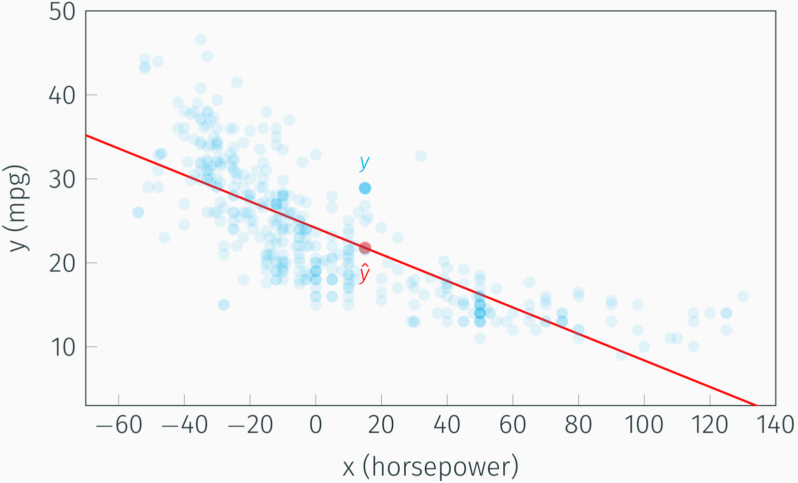
Linear regression (via ordinary least squares)



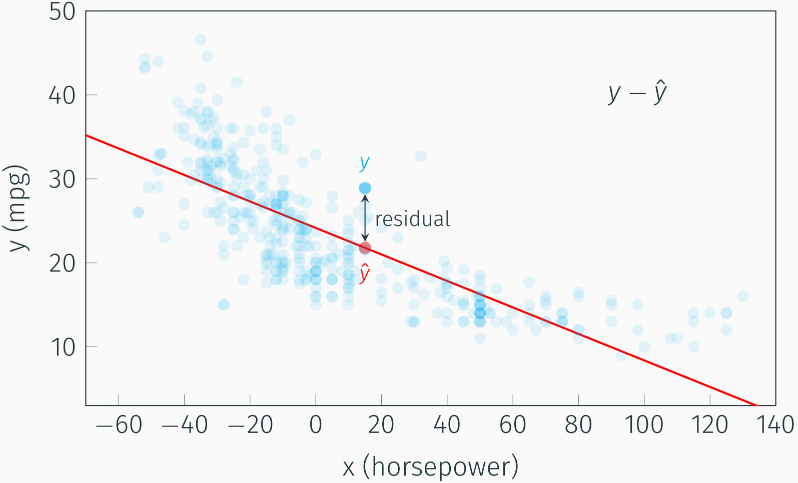
Linear regression (via ordinary least squares)



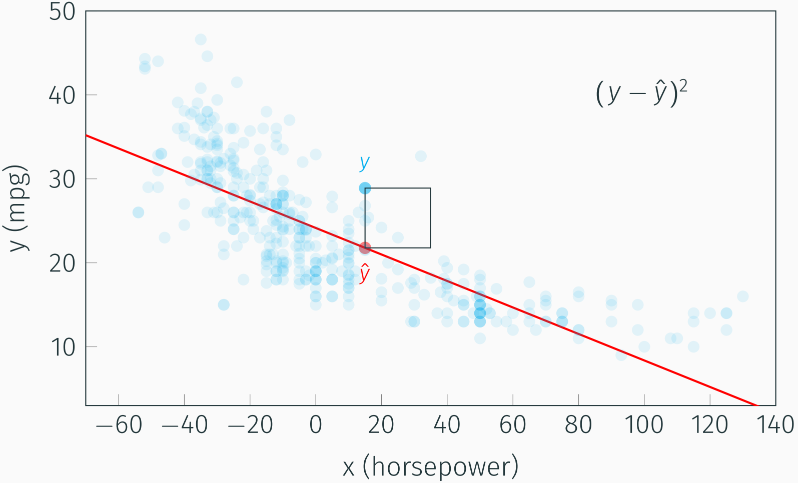
Linear regression (via ordinary least squares)



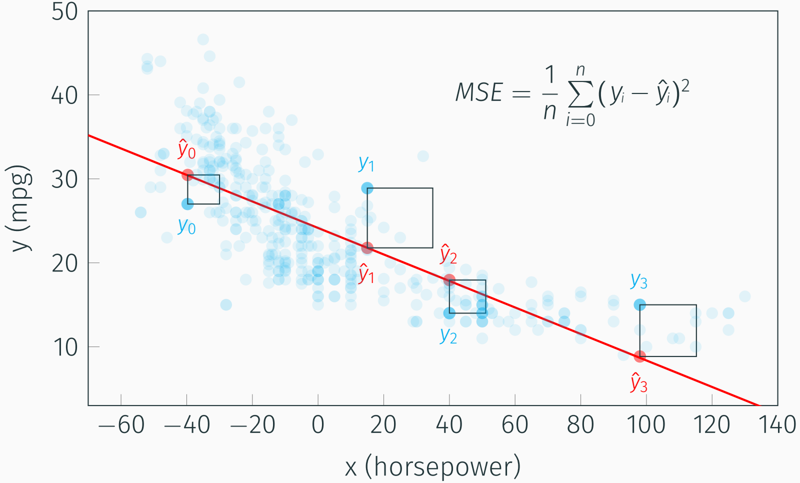
Linear regression (via ordinary least squares)



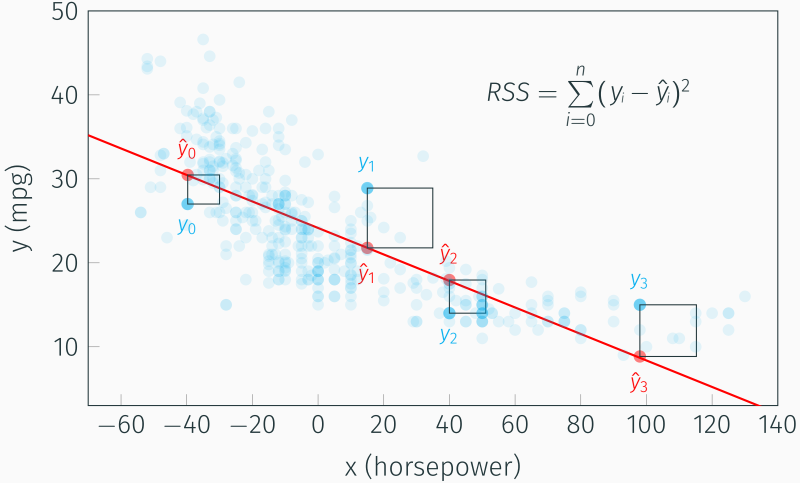
Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

Linear regression: Models the relationship between input x and output y by finding the linear model $\hat{y} = \beta_0 + \beta_1 x$ that minimizes the residual sum of squares (RSS).

- β_0 refers to the intercept (or offset) of the model
- β_1 refers to the slope of the model

Fitting a linear regression model

$$\hat{y} = \beta_0 + \beta_1 x$$

Fitting a linear regression model

$$\hat{y} = \beta_0 + \beta_1 x$$

Fitting a linear regression model

$$\hat{y} = \beta_0 + \beta_1 x$$
$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Fitting a linear regression model

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 + \beta_1 x_i)^2$$

Fitting a linear regression model

Fitting a linear regression model

Fitting a linear regression model

Fitting a linear regression model

Fitting a linear regression model

Fitting a linear regression model

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Multivariate linear regression

Linear regression: The true workhorse of machine learning

- Models the relationship between (either singular or mulitple) inputs X and (a continuous) output y as a linear function
  - Inputs can be both continuous and categorical
- A strict parametric form limits the expressivity of the model
  - More advanced terms can be explicitly added
  - The strictness allows for extended functionality, such as computing confidence intervals
  - **Makes the model human interpretable**