PSY9511: Seminar 4

The basics of regression and classification

Esten H. Leonardsen

29.05.2024

UNIVERSITETET
I OSLO

Recap

What is statistical learning?

Recap

What is statistical learning?

Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.

Recap

What is statistical learning?

Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.

Predictive view: Finding a function $\hat{f}(X)$ that, when given a new set of inputs X allows us to predict an output y.

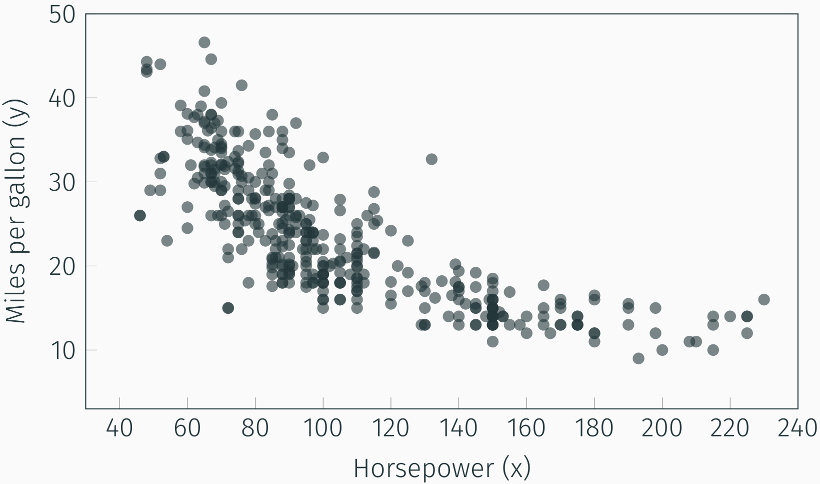
Recap

What is statistical learning?

Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.

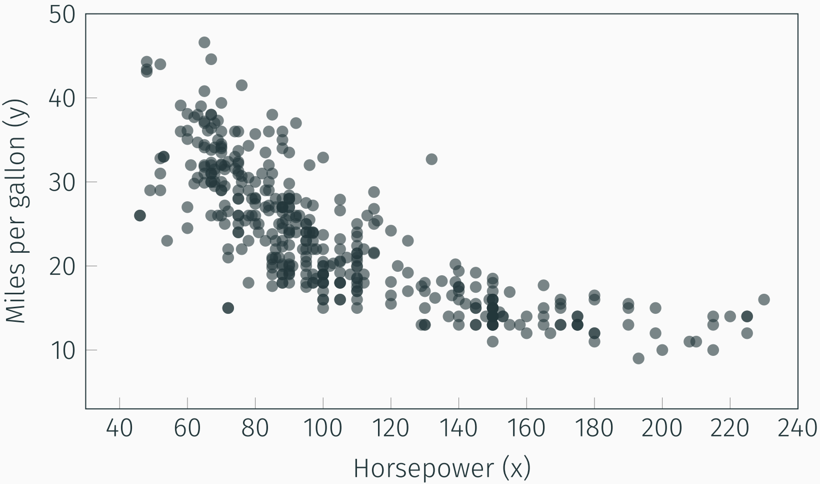
Predictive view: Finding a function $\hat{f}(X)$ that, when given a new set of inputs X allows us to predict an output y.

Recap



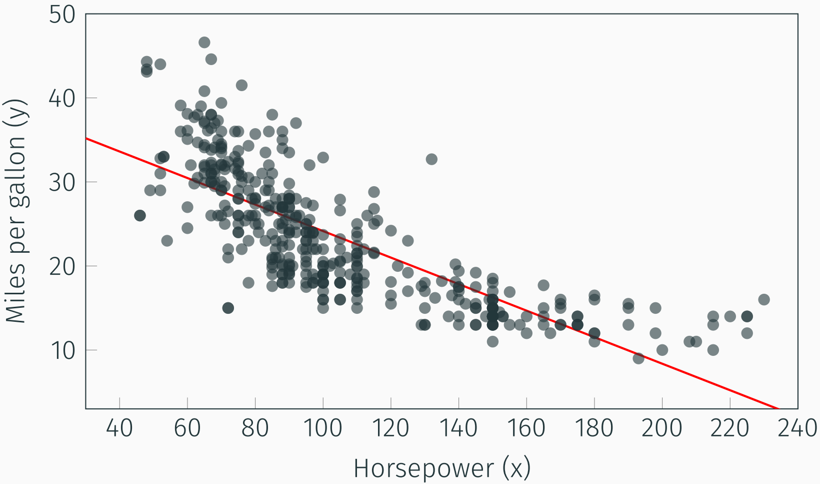
Miles per gallon (y)
50
40
30
20
10
40
60
80
100
120
140
160
180
200
220
240
Horsepower (x)

Recap



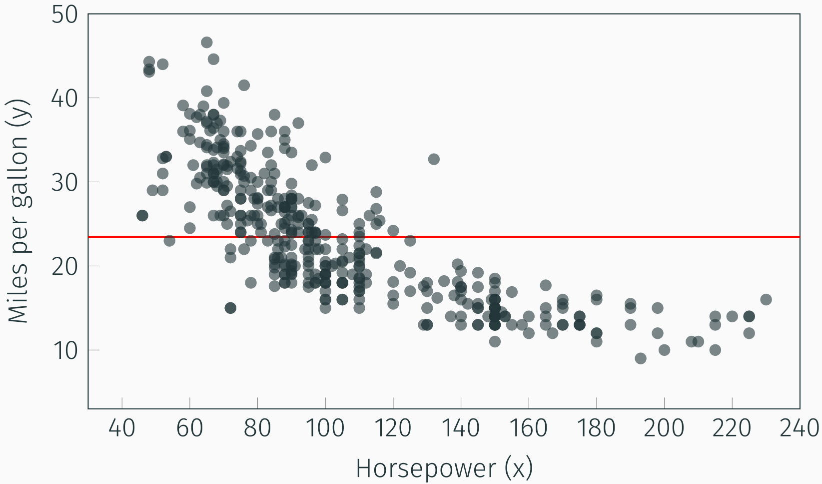
$$\hat{y} = \hat{f}(x)$$

Recap



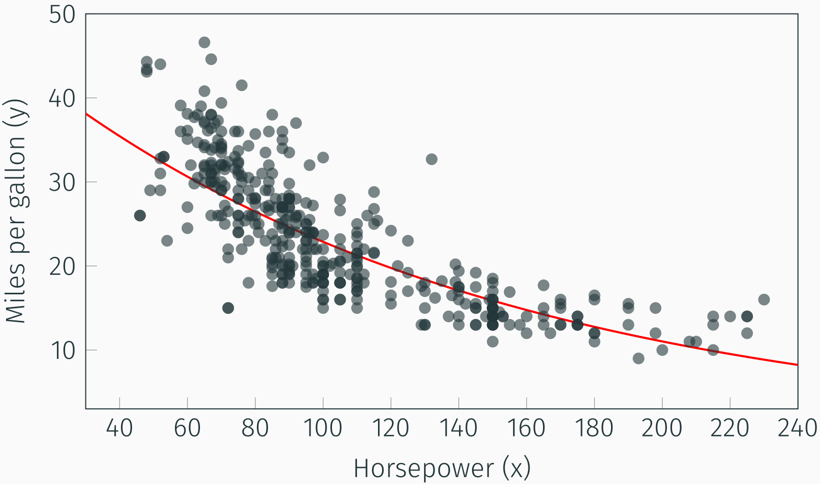
$$\hat{y} = 39.93 - 0.1578x$$

Recap



$\hat{y} = 23.44$

Recap



$$\hat{y} = e^{3.86-0.0073x}$$

Recap

Bias-variance tradeoff

Outline

Plan for the day:

- Different types of outputs y: Regression vs classification
- Linear regression: Restricting the scope of $\hat{f}(X)$
  - Live coding
- k-Nearest Neighbours
- Logistic regression: Extending linear regression to classification
  - Live coding
- Generative models

Plan for future lectures:

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

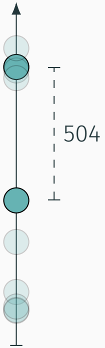
Regression vs. classification

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

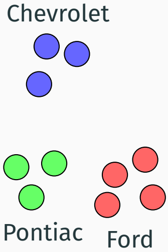
Regression vs. classification



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

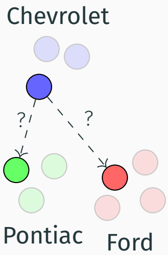
Regression vs. classification

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

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

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

Regression vs. classification

Regression:

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

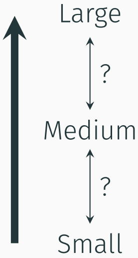
Regression vs. classification

Large

Medium

Small

Regression vs. classification

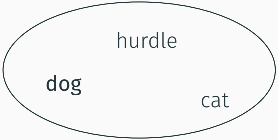


Regression vs. classification

The quick brown fox jumps over the lazy ______

Regression vs. classification

The quick brown fox jumps over the lazy ______



Classification

Regression vs. classification

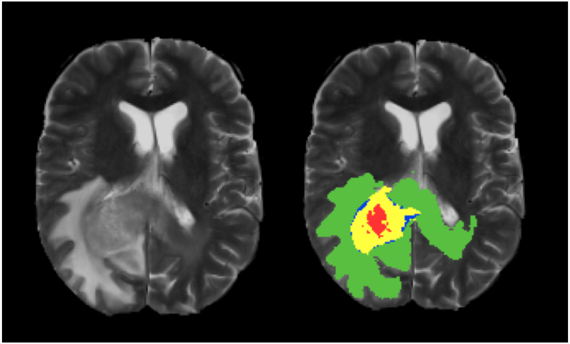
"Students taking a machine learning class"

Regression vs. classification

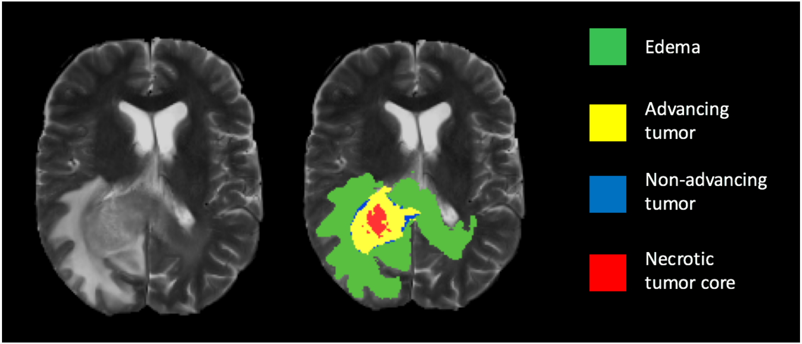
”Students taking a machine learning class” → →

(128, 192, 64)

Regression vs. classification



Regression vs. classification



Regression vs. classification

Regression vs. classification

Regression vs. classification

Speech generation

Regression vs. classification

Speech generation

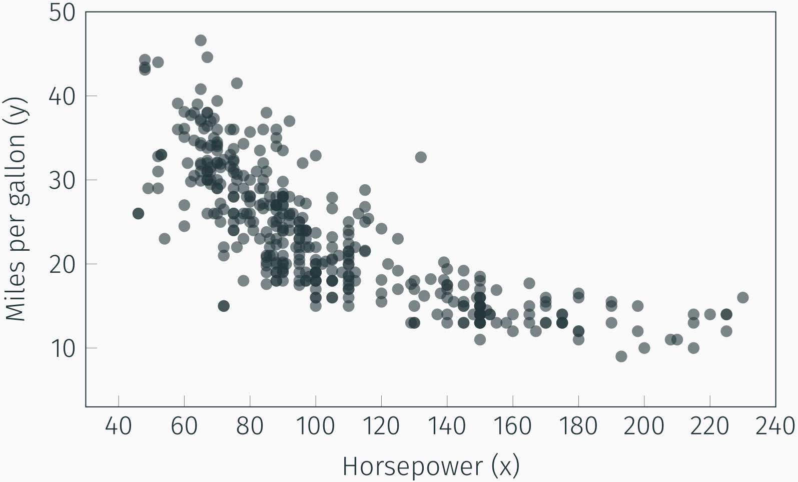
Regression vs. classification

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

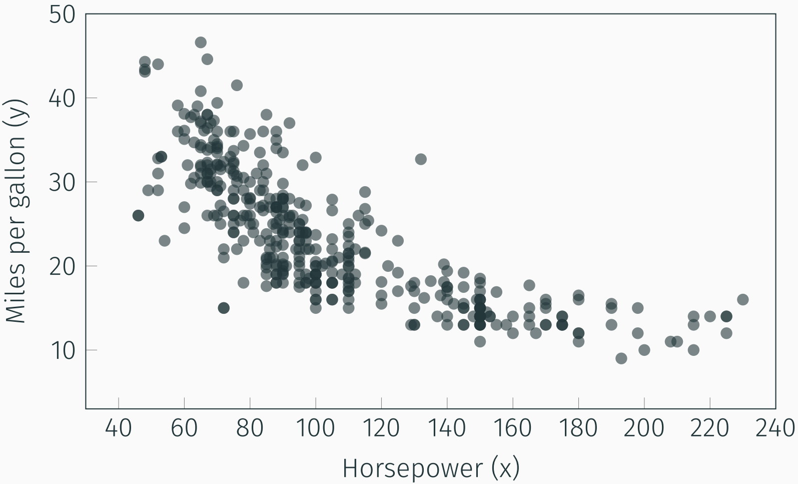
- Problems with quantitative outputs are solved via regression, often by minimizing the mean squared error.
- Problems with qualitative outputs are solved by classification, often by maximizing accuracy.
- Ordinal regression falls between the two, with disjunct classes that can be ordered.
- A variety 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)

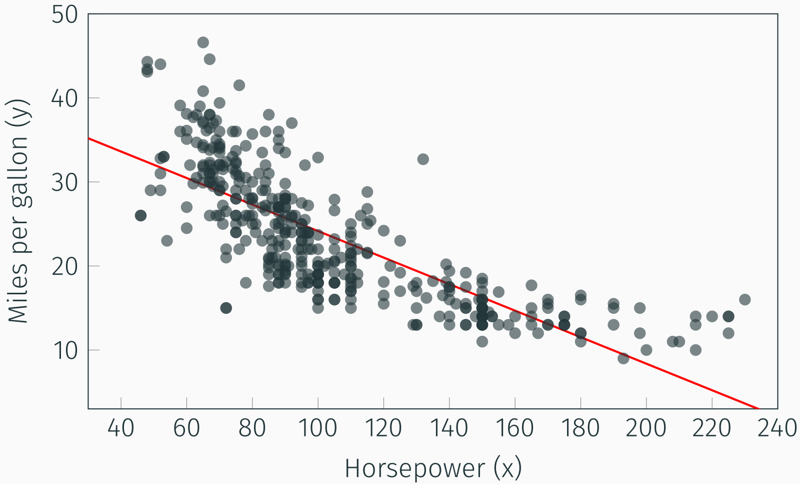


Linear regression (via ordinary least squares)



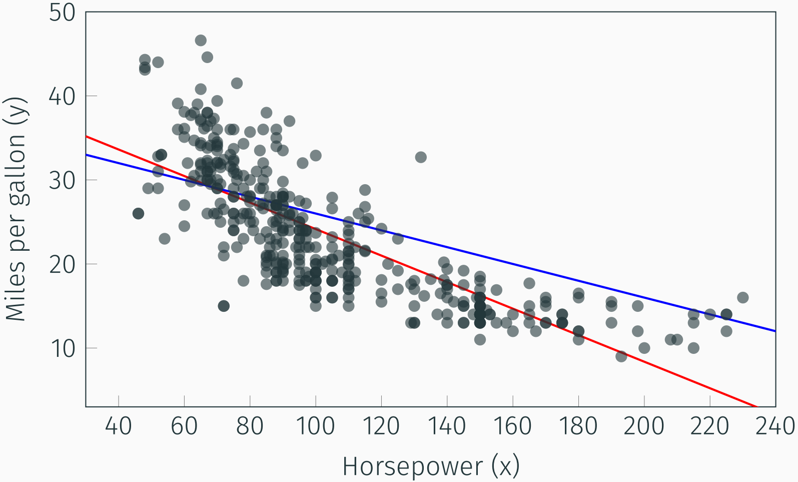
$$\hat{y} = \beta_0 - \beta_1 x$$

Linear regression (via ordinary least squares)



$$\hat{y} = 39.93 - 0.1578x$$

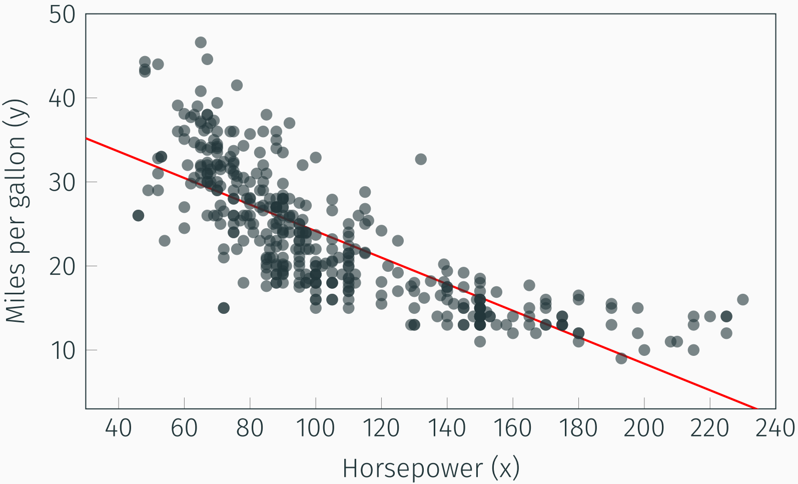
Linear regression (via ordinary least squares)



$$\hat{y} = 39.93 - 0.1578x$$

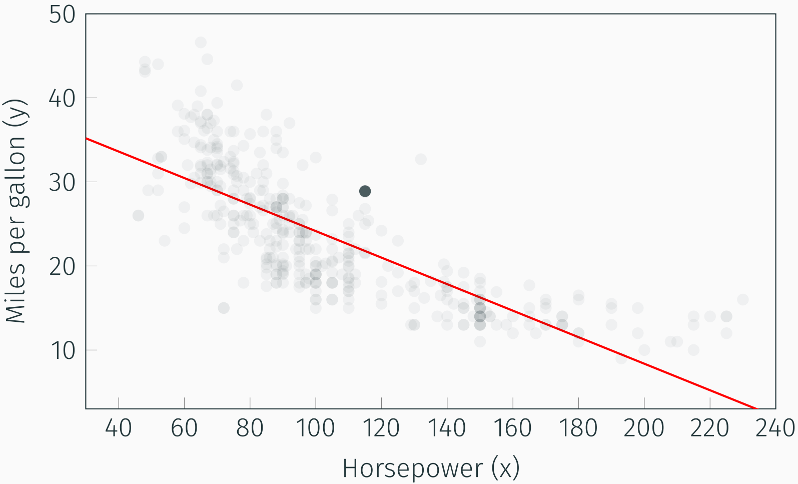
$$\hat{y} = 36 - 0.1x$$

Linear regression (via ordinary least squares)



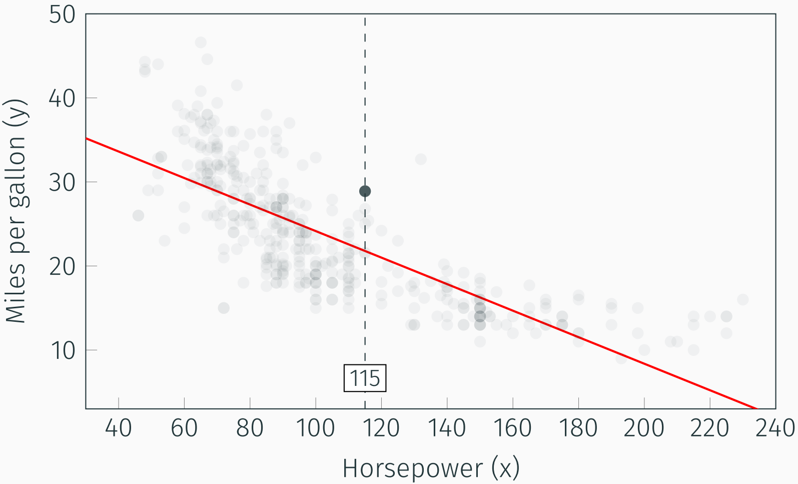
$$\hat{y} = 39.93 - 0.1578x$$

Linear regression (via ordinary least squares)



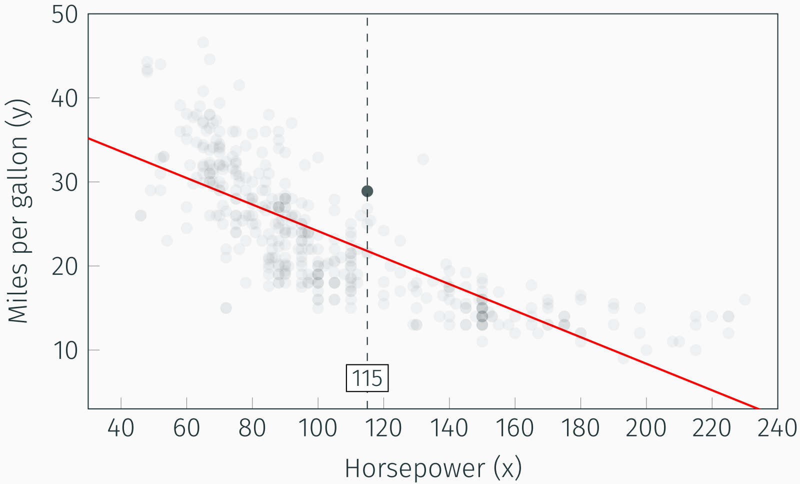
$$\hat{y} = 39.93 - 0.1578x$$

Linear regression (via ordinary least squares)



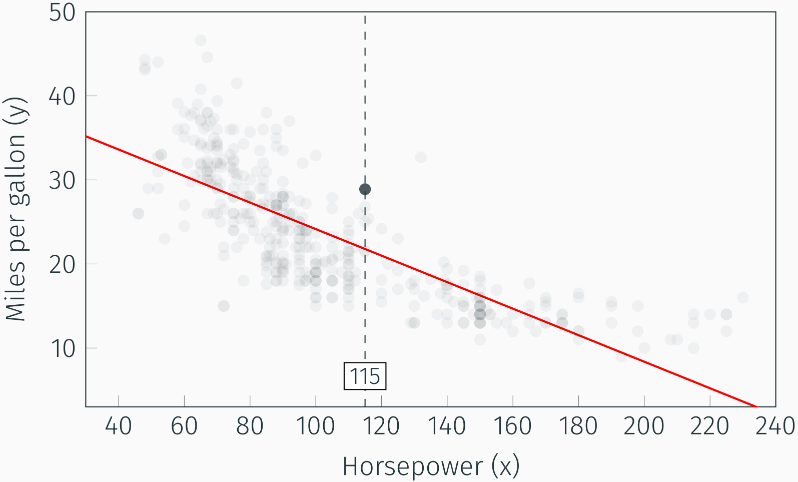
$$\hat{y} = 39.93 - 0.1578x$$

Linear regression (via ordinary least squares)



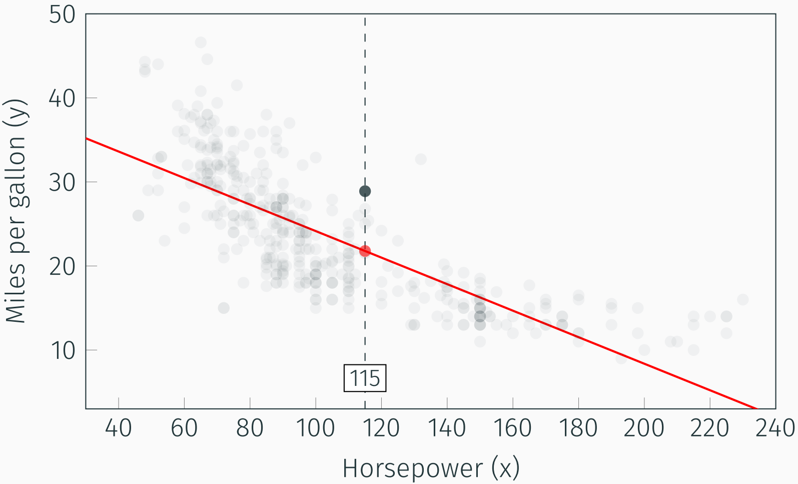
$$\hat{y} = 39.93 - 0.1578 * 115$$

Linear regression (via ordinary least squares)



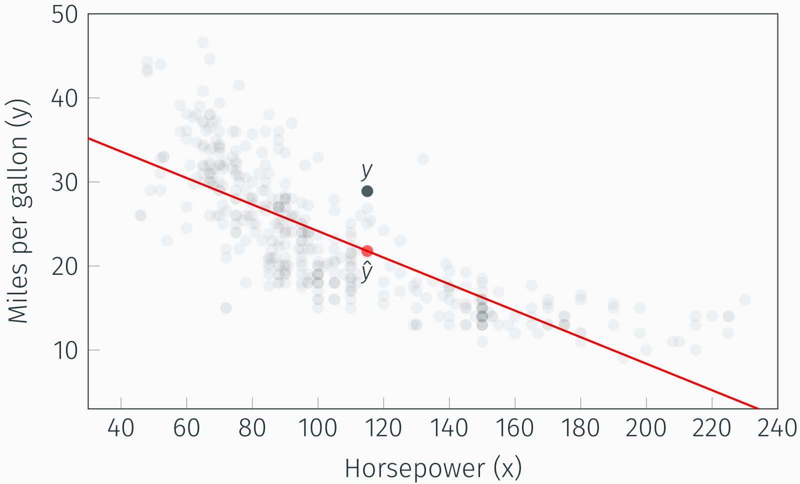
$$21.78 = 39.93 - 0.1578 * 115$$

Linear regression (via ordinary least squares)

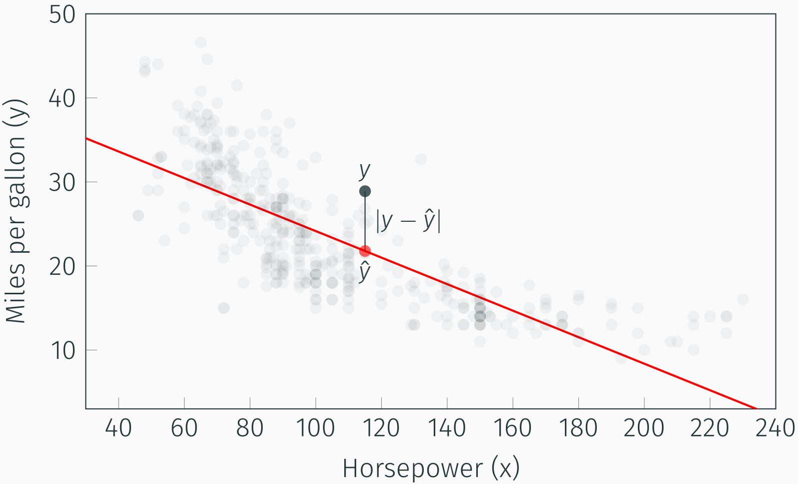


$$21.78 = 39.93 - 0.1578 * 115$$

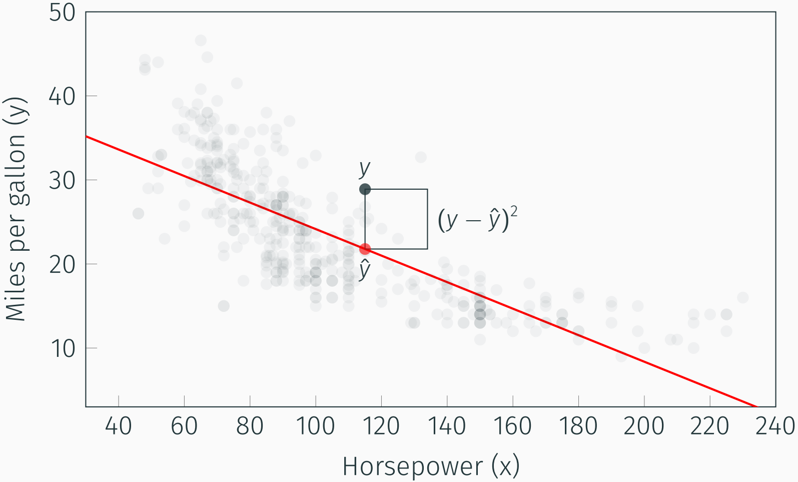
Linear regression (via ordinary least squares)



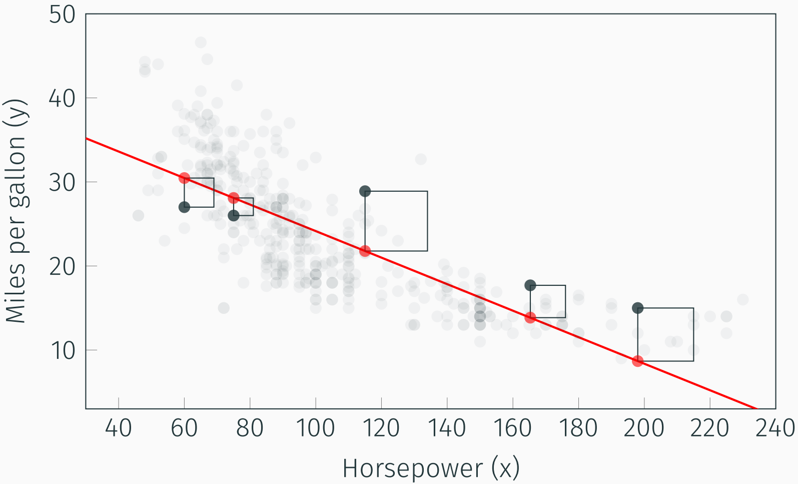
Linear regression (via ordinary least squares)



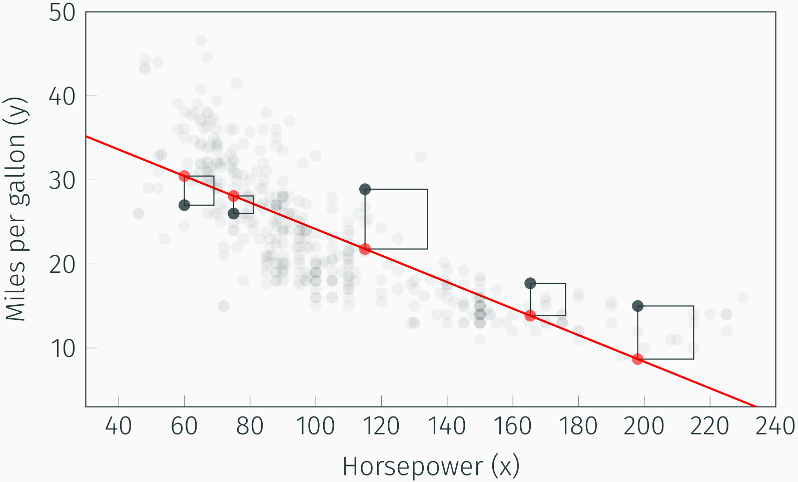
Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

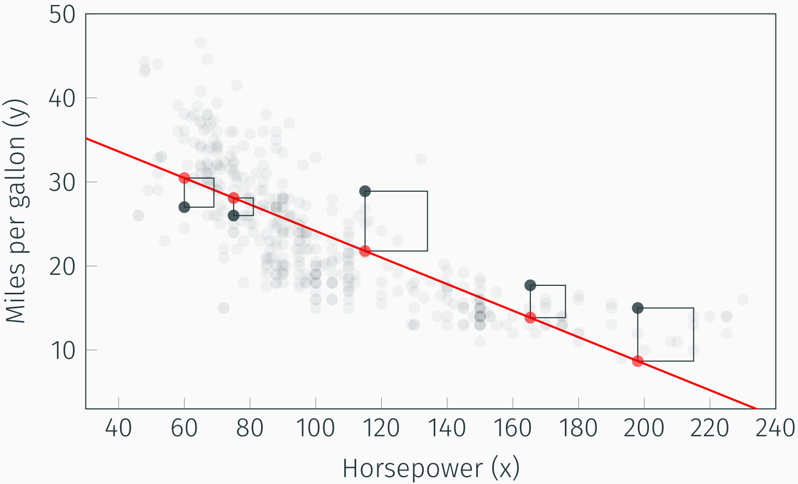


Linear regression (via ordinary least squares)



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

Linear regression (via ordinary least squares)



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

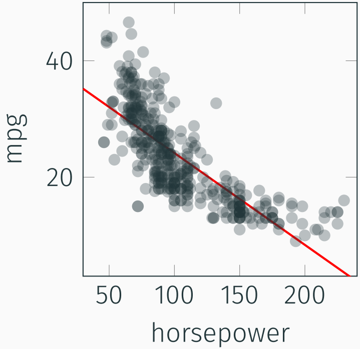
Linear regression (via ordinary least squares)

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

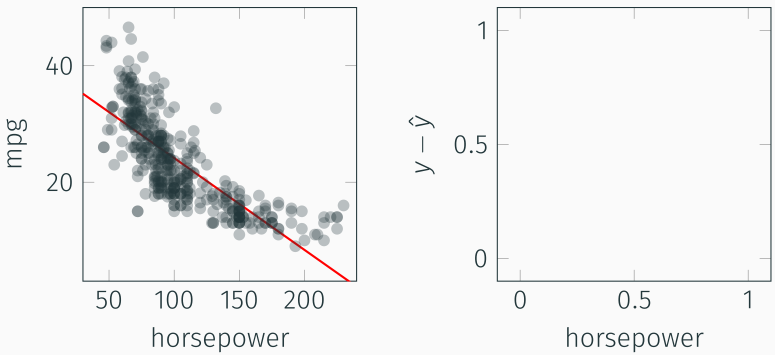
Linear regression (via ordinary least squares)

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

Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

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

Linear regression (via ordinary least squares)

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

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

Linear regression (via ordinary least squares)

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

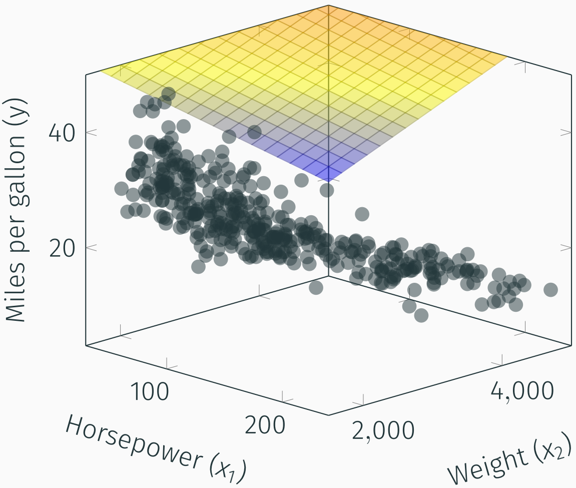
Linear regression (via ordinary least squares)

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

Linear regression (via ordinary least squares)

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

mpg	manufacturer
36	Chevrolet
15	Ford
25	Chevrolet
26	Chevrolet
17	Ford
15	Ford
32	Chevrolet
14	Ford
14	Ford
28	Chevrolet

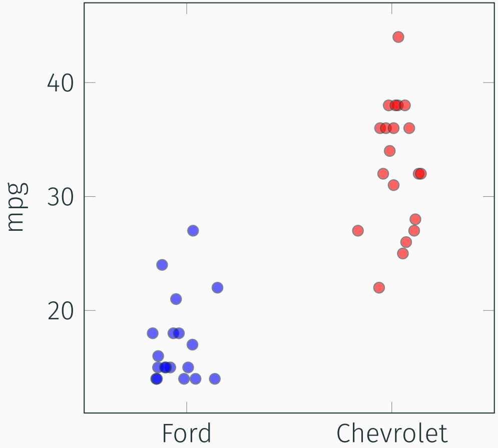
$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{manufacturer}$$

Linear regression (via ordinary least squares)

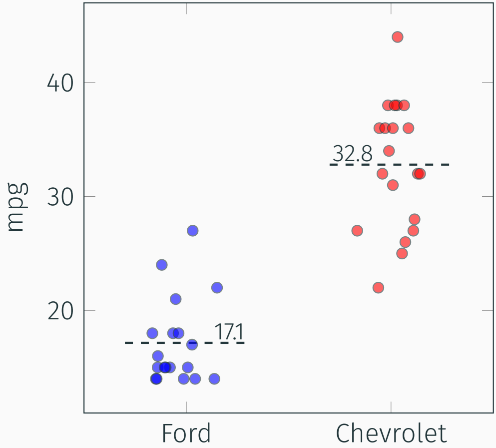
mpg	manufacturer	chevrolet
36	Chevrolet	1
15	Ford	0
25	Chevrolet	1
26	Chevrolet	1
17	Ford	0
15	Ford	0
32	Chevrolet	1
14	Ford	0
14	Ford	0
28	Chevrolet	1

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet}$$

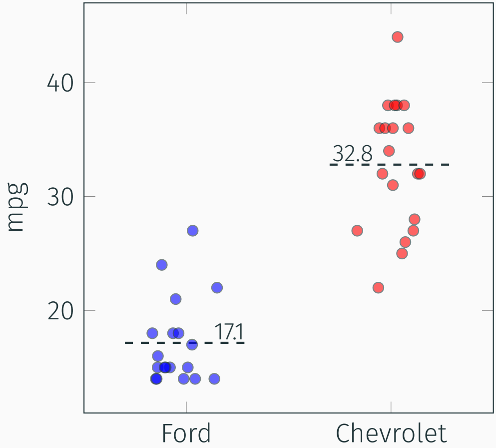
Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Blackboard!

Linear regression (via ordinary least squares)

mpg	manufacturer
36	Chevrolet
15	Ford
25	Chevrolet
26	Pontiac
17	Ford
15	Ford
32	Pontiac
14	Ford
14	Pontiac
28	Chevrolet

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{manufacturer}$$

Linear regression (via ordinary least squares)

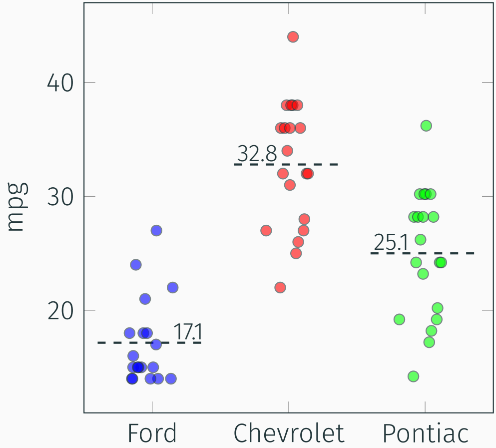
mpg	manufacturer	chevrolet	pontiac
36	Chevrolet	1	0
15	Ford	0	0
25	Chevrolet	1	0
26	Pontiac	0	1
17	Ford	0	0
15	Ford	0	0
32	Pontiac	0	1
14	Ford	0	0
14	Pontiac	0	1
28	Chevrolet	1	0

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet} + \beta_2 \times \text{pontiac}$$

Linear regression (via ordinary least squares)

Python dummy encoding

Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

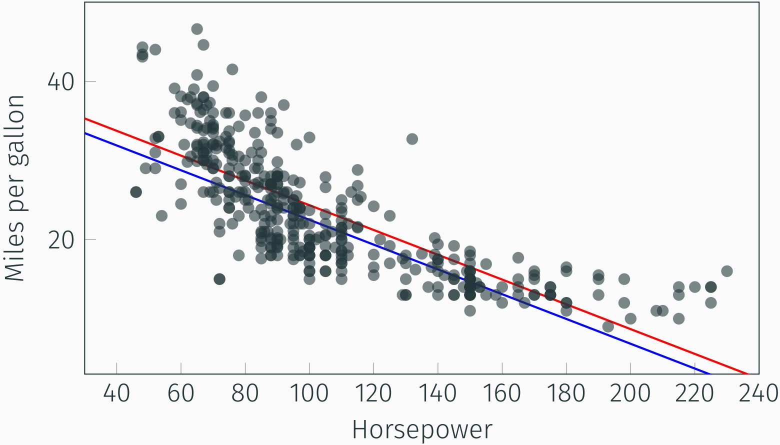
mpg	chevrolet	horsepower
36	1	130
15	0	165
25	1	150
26	1	150
17	0	140
15	0	198
32	1	220
14	0	215
14	0	225
28	1	212

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet} + \beta_2 \times \text{horsepower}$$

Linear regression (via ordinary least squares)

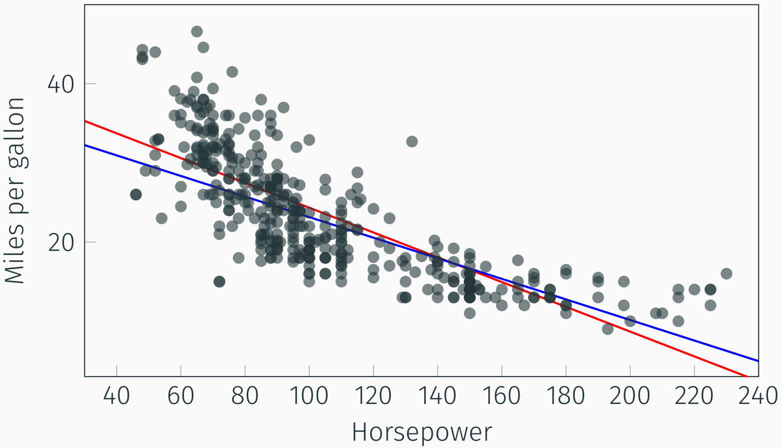
$$\widehat{mpg} = \begin{cases} \beta_0 + \beta_1 + \beta_2 \times \text{horsepower} & \text{if chevrolet} \\ \beta_0 + \beta_2 \times \text{horsepower} & \text{else} \end{cases}$$

Linear regression (via ordinary least squares)

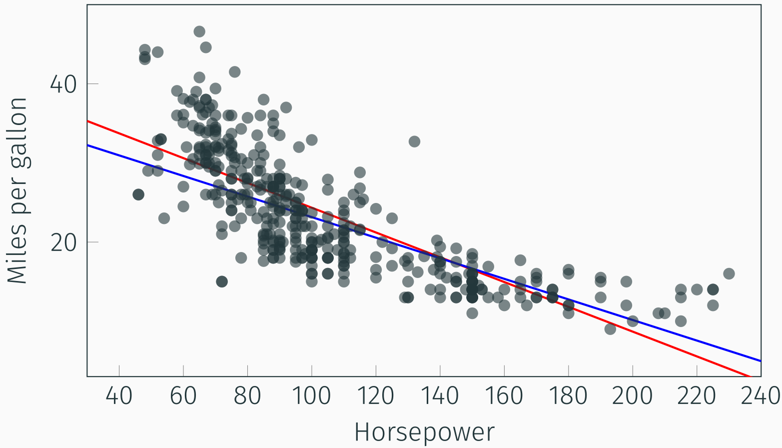


$$\widehat{mpg} = \begin{cases} \beta_0 + \beta_1 + \beta_2 \times \text{horsepower} & \text{if chevrolet} \\ \beta_0 + \beta_2 \times \text{horsepower} & \text{else} \end{cases}$$

Linear regression (via ordinary least squares)

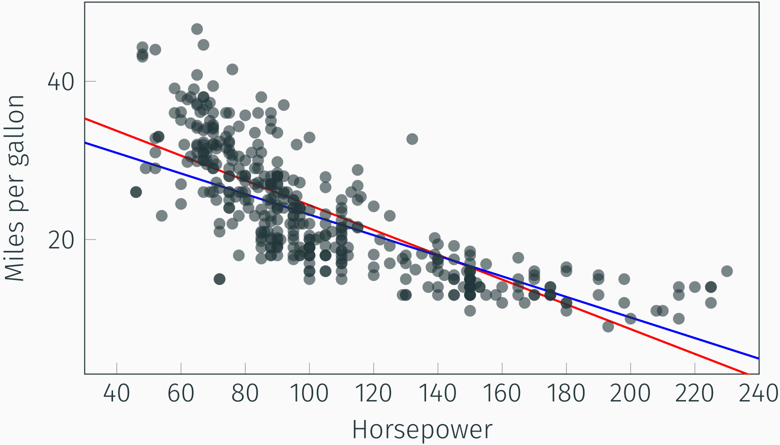


Linear regression (via ordinary least squares)



$$\widehat{mpg} = \beta_0 + \beta_1 \times \text{chevrolet} + \beta_2 \times \text{horsepower}$$

Linear regression (via ordinary least squares)



$$\widehat{mpg} = \beta_0 + \beta_1 \times \text{chevrolet} + \beta_2 \times \text{horsepower}$$
$$+\beta_3 \times \text{chevrolet} \times \text{horsepower}$$

Confidence intervals

Linear regression (via ordinary least squares)

Evaluation

Linear regression (via ordinary least squares)

Live coding

Linear regression (via ordinary least squares)

Linear regression:

$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

K-Nearest Neighbours:

$$\hat{f}(X) = \frac{1}{K} \sum_{x_i \in \mathcal{N}} y_i$$

k nearest neighbours

Linear regression:

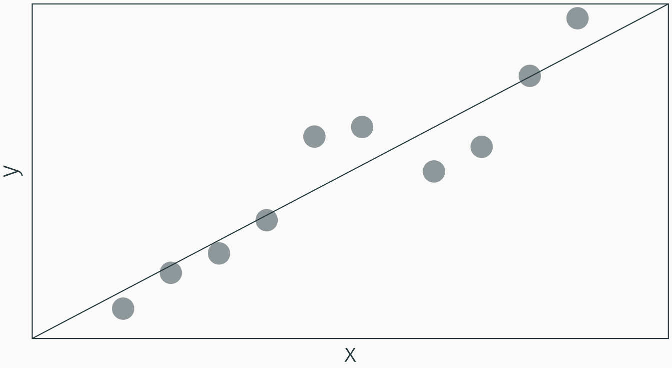
$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

Blackboard!

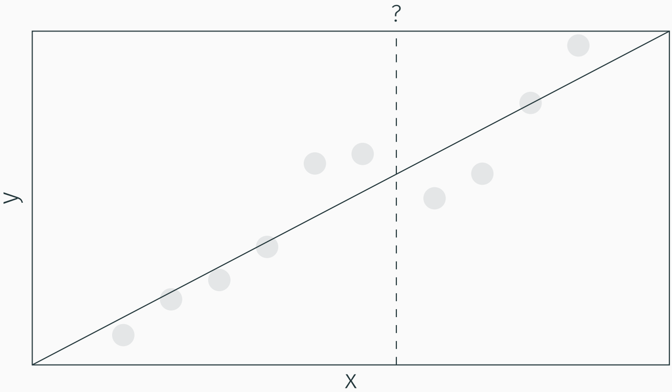
K-Nearest Neighbours:

$$\hat{f}(X) = \frac{1}{K} \sum_{x_i \in \mathcal{N}} y_i$$

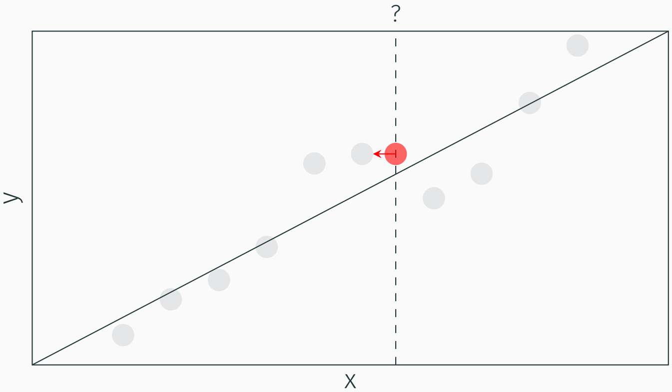
k nearest neighbours



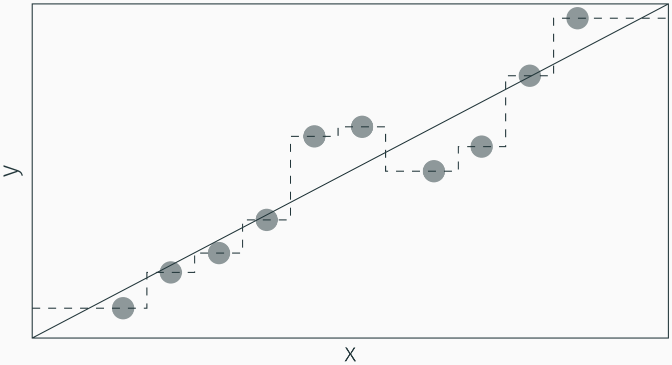
k nearest neighbours



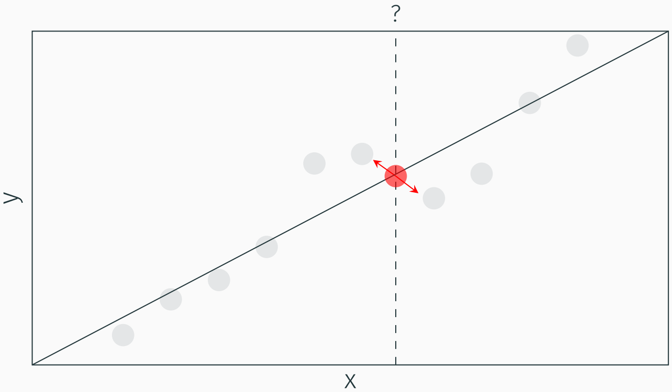
k nearest neighbours



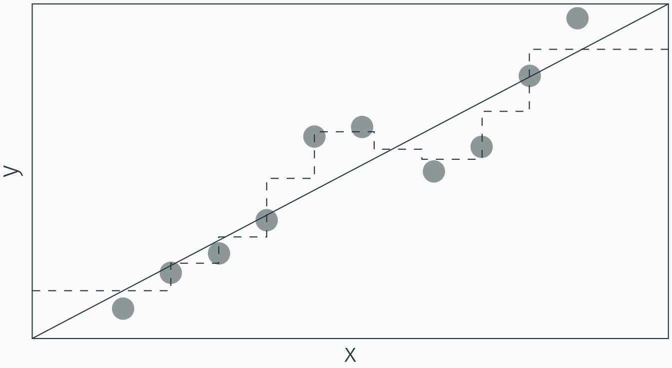
k nearest neighbours



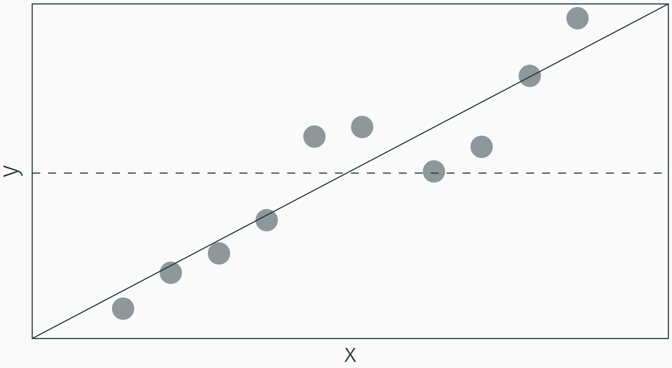
k nearest neighbours



k nearest neighbours



k nearest neighbours



k nearest neighbours

Curse of dimensionality

Logistic regression

mpg	manufacturer	chevrolet
36	Chevrolet	1
15	Ford	0
25	Chevrolet	1
26	Chevrolet	1
17	Ford	0
15	Ford	0
32	Chevrolet	1
14	Ford	0
14	Ford	0
28	Chevrolet	1

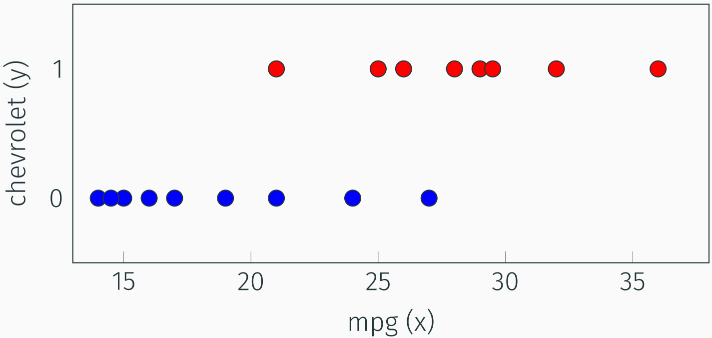
$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet}$$

Logistic regression

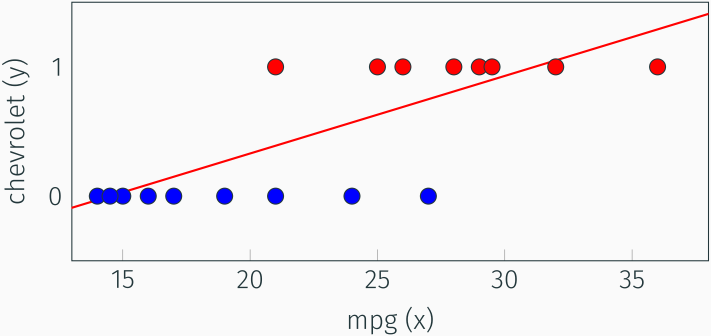
mpg	manufacturer	chevrolet
36	Chevrolet	1
15	Ford	0
25	Chevrolet	1
26	Chevrolet	1
17	Ford	0
15	Ford	0
32	Chevrolet	1
14	Ford	0
14	Ford	0
28	Chevrolet	1

$$\widehat{\text{chevrolet}} = \beta_0 + \beta_1 \times \text{mpg}$$

Logistic regression

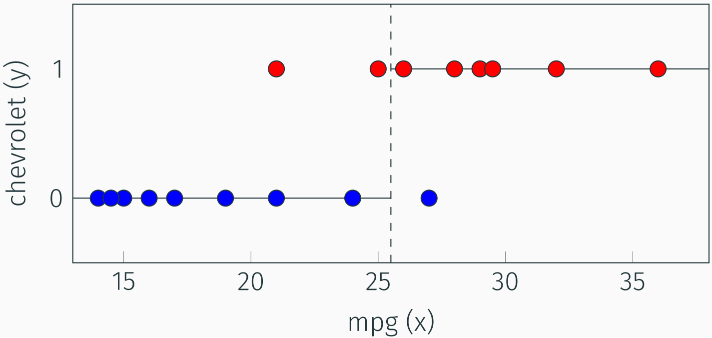


Logistic regression

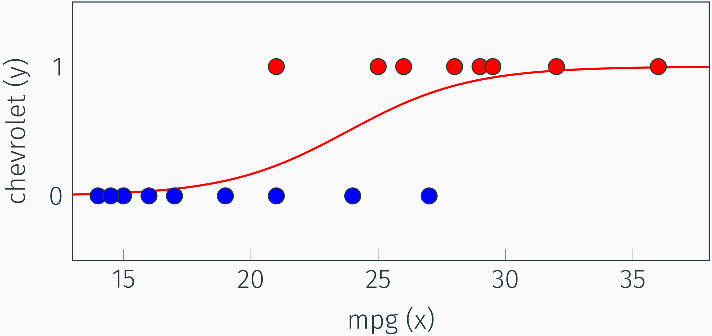


$$\widehat{\text{chevrolet}} = -0.87 + 0.06 \times \text{mpg}$$

Logistic regression



Logistic regression



$$\widehat{\text{chevrolet}} = \frac{e^{-10.22+0.42\times \text{mpg}}}{1+e^{-10.22+0.42\times \text{mpg}}}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

1. Given that $0 \leq \hat{y} \leq 1$, we can interpret it as a probability: $\hat{y} = Pr(Y = 1|X)$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

1. Given that $0 \leq \hat{y} \leq 1$, we can interpret it as a probability: $\hat{y} = Pr(Y = 1|X)$
2. The log-odds of belonging to the positive class is linear with respect to the coefficients: $log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 x$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

1. Given that $0 \leq \hat{y} \leq 1$, we can interpret it as a probability: $\hat{y} = Pr(Y = 1|X)$

2. The log-odds of belonging to the positive class is linear with respect to the coefficients: $log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 x$

3. Multiclass

Generative models