PSY9511: Seminar 4

Model selection, validation and testing

Esten H. Leonardsen

23.09.24

UNIVERSITETET
I OSLO

Outline

1. Assignment 3
2. Loss functions and performance metrics
3. Strategies for model evaluation
    - Training and validation split
    - (Stratification)
    - (Leave-one-out cross-validation)
    - Cross-validation
    - Bootstrap
    - Model comparison
4. Strategies for model selection **and** evaluation
    - Train/validation/test split
    - Nested cross-validation

Assigment 3

Assignment 3

Loss functions and performance metrics

Loss functions and performance metrics

Commonalities

- Allows us to evaluate the performance of a model
- Typically on the form $f(y, \hat{y})$

Loss functions

- Tailored specifically for mathematical optimization of models

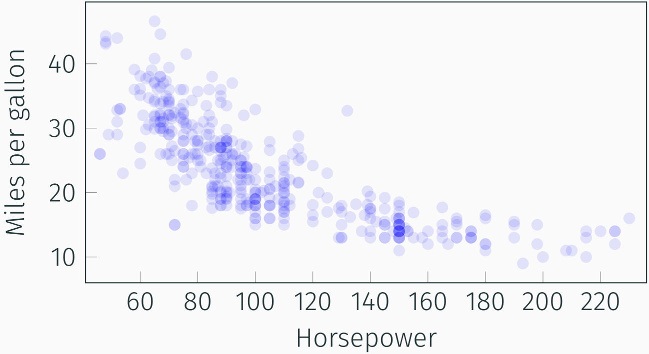
Performance metrics

- Tailored specifically for interpretation of model performance by humans

Loss functions and performance metrics

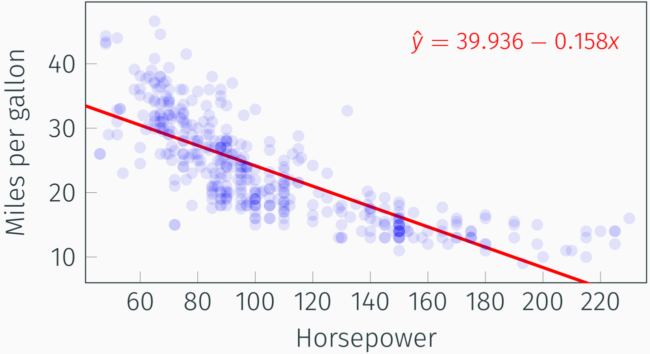
$$\text{mse}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

Loss functions and performance metrics



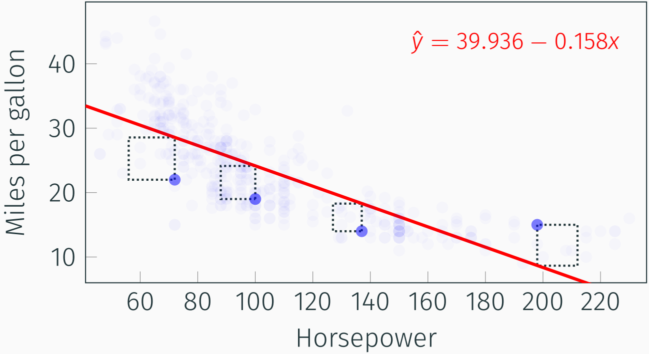
$$\mathrm{mse}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

Loss functions and performance metrics



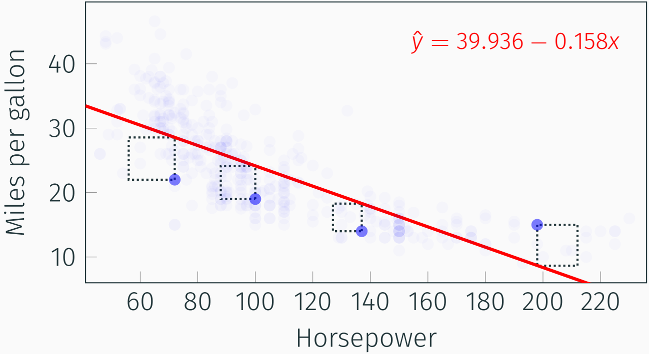
$$\text{mse}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2$$

Loss functions and performance metrics



$$\mathrm{mse}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2$$

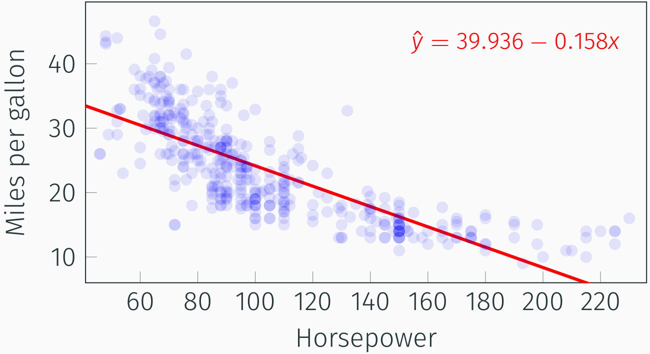
Loss functions and performance metrics



$$\mathrm{mse}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

$$= 23.94$$

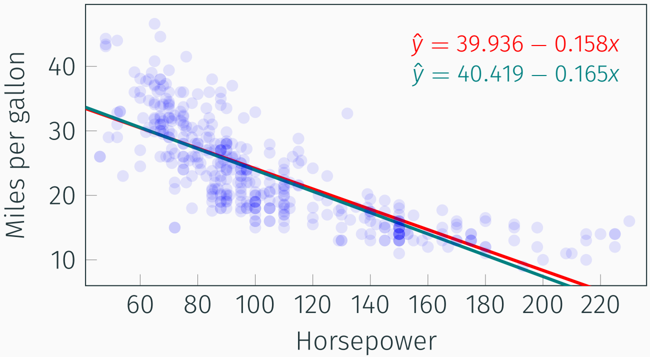
Loss functions and performance metrics



$$\mathrm{mse}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2$$

$$\mathrm{mae}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n} |y_i - \hat{y}_i|$$

Loss functions and performance metrics



$$\text{mse}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

$$\text{mae}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n}|y_i - \hat{y}_i|$$

Loss functions and performance metrics

Loss functions

- Different loss functions measures different properties of the model fit
- Optimizing for them gives different parameter estimates

Loss functions and performance metrics

Tolerance-based accuracy:

A prediction is considered correct if it is within a predefined margin of error from the true value

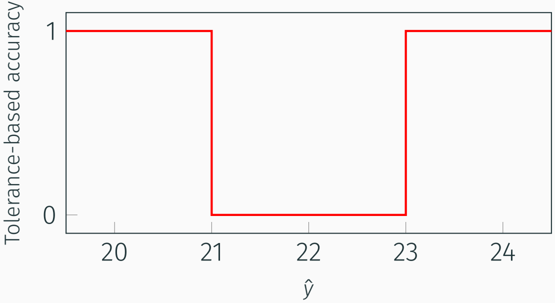
$$\text{accuracy*}(y, \hat{y}) = \begin{cases} 0 & \text{if } |y - \hat{y}| < \text{tol} \\ 1 & else \end{cases}$$

Loss functions and performance metrics

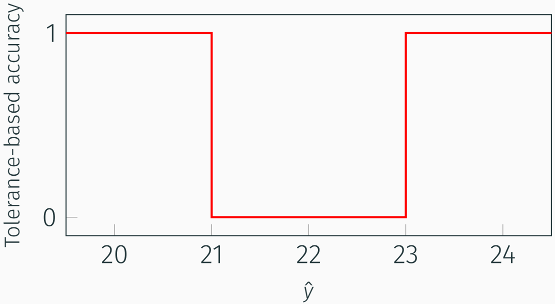
mpg	horsepower
22	72

Loss functions and performance metrics

mpg	horsepower
22	72



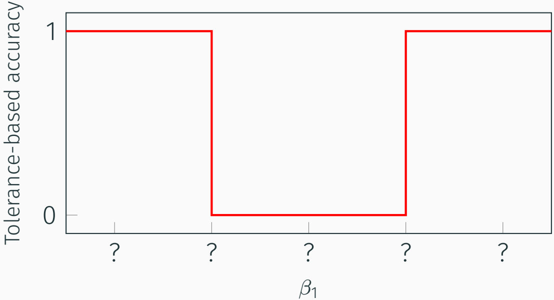
mpg	horsepower
22	72



$$\hat{y} = \beta_0 + \beta_1 \times \text{horsepower}$$

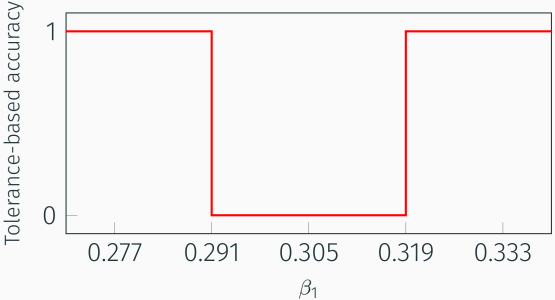
Loss functions and performance metrics

mpg	horsepower
22	72



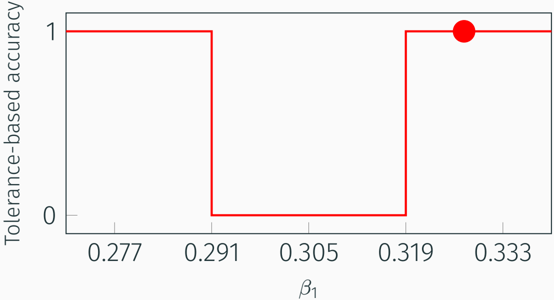
$$\hat{y} = \beta_0 + \beta_1 \times \text{horsepower}$$

Loss functions and performance metrics



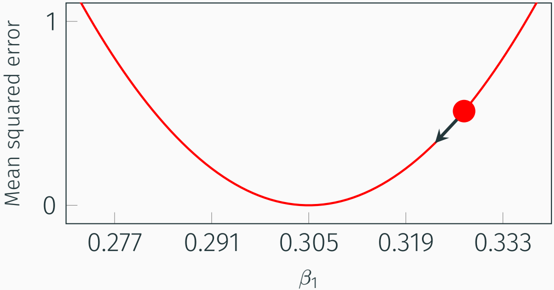
$$\hat{y} = 0 + \beta_1 \times \text{horsepower}$$

Loss functions and performance metrics



$$\hat{y} = 0 + 0.33 \times \text{horsepower}$$

Loss functions and performance metrics



$$\hat{y} = 0 + 0.33 \times \text{horsepower}$$

Loss functions

- Different loss functions measures different properties of the model fit
- Optimizing for them gives different parameter estimates
- Must be differentiable to allow for mathematical optimization

Loss functions and performance metrics

$$\text{mse}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

OR

$$\text{mae}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n}|y_i - \hat{y}_i|$$

Performance metrics: Regression

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

Mean squared error (MSE)

+ Can be used as a loss function
+ Widely used
+ Intuitive
+ Penalizes large errors

? Interpretation

- Depends on scale

$$\sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}$$

Root mean squared error (RMSE)

+ Can be used as a loss function
+ Intuitive
+ Penalizes large errors
+ More interpretable than MSE, total loss $\approx$ individual loss
- Depends on scale

Performance metrics: Regression

$$\frac{1}{n}\sum_{i=0}^{n} |y_i - \hat{y}_i|$$

Mean absolute error (MAE)

- + Can be used as a loss function
- + More interpretable than MSE/RMSE, total loss = average error
- – Feels a bit off
- – Depends on scale

Performance metrics: Regression

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

Pearson correlation coefficient (r)

+ Scale independent
? Captures linear correlation
- Should not be used as a loss function
- Does not care about whether the predictions are close to the true values

Performance metrics: Regression

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

Proportion of variance explained (r^2)

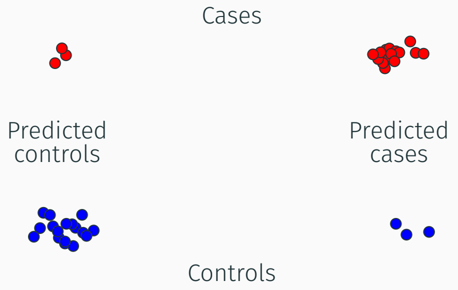
+ Scale independent
+ Interpretable

? Captures linear correlation

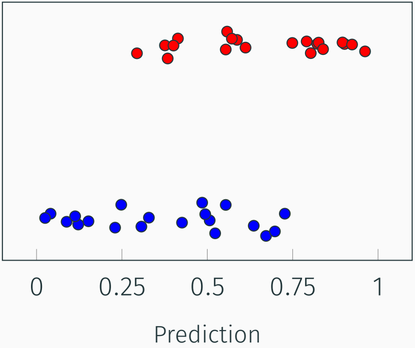
- Should not be used as a loss function
- Does not care about whether the predictions are close to the true values

Performance metrics: Binary classification

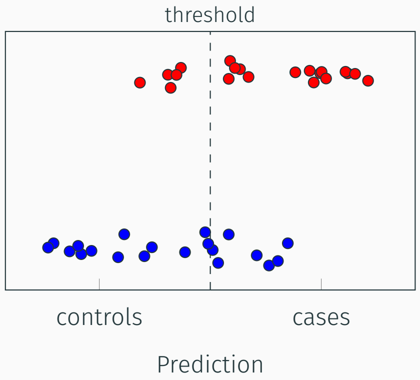
Performance metrics: Binary classification



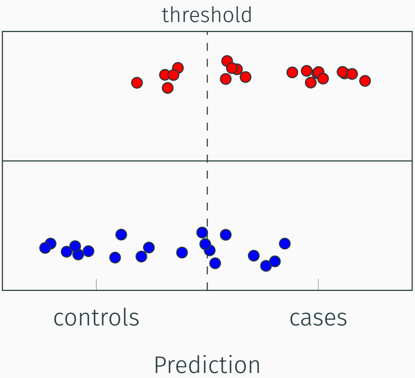
Performance metrics: Binary classification



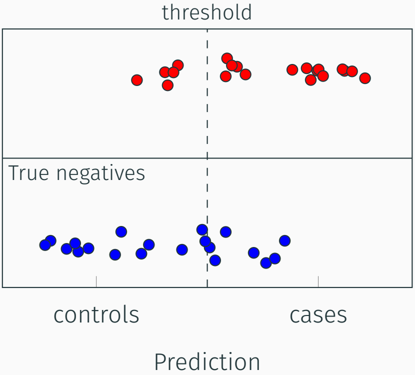
Performance metrics: Binary classification



Performance metrics: Binary classification

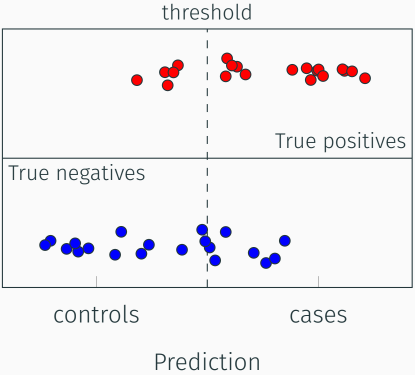


Performance metrics: Binary classification



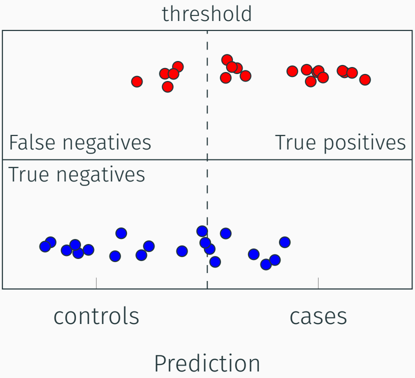
TN	

Performance metrics: Binary classification



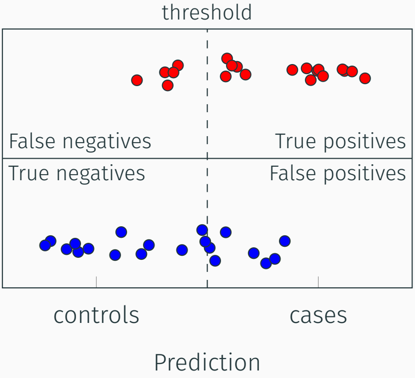
TN	
	TP

Performance metrics: Binary classification



TN	
FN	TP

Performance metrics: Binary classification



TN	FP
FN	TP