Q & A: Linear Classification

Lecture Series „Machine Learning“

Niels Landwehr

Research Group „Data Science“
Institute of Computer Science
University of Hildesheim

Quiz: Logistic Regression

- Assume we have learned a binary logistic regression model $f:\mathbb{R}^2 \to [0,1]$ of the form

  $$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(\theta_1 x_1 + \theta_2 x_2)$$

  by maximizing the conditional log-likelihood on a data set

  $$\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$$

  as explained in the lecture. Assume the learned parameter vector is given by

  $$\boldsymbol{\theta}^* = (0.5, -1)$$

- **Question 1**: For the data point $\mathbf{x}_0 = (1,1)$, what is $p(y=1 \mid \mathbf{x}_0, \boldsymbol{\theta}^*)$?

Quiz: Logistic Regression

- Assume we have learned a binary logistic regression model $f:\mathbb{R}^2 \rightarrow [0,1]$ of the form

  $$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(\theta_1 x_1 + \theta_2 x_2)$$

  by maximizing the conditional log-likelihood on a data set

  $$\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$$

  as explained in the lecture. Assume the learned parameter vector is given by

  $$\boldsymbol{\theta}^* = (0.5, -1)$$

- **Question 1**: For the data point $\mathbf{x}_0 = (1,1)$, what is $p(y=1 \mid \mathbf{x}_0, \boldsymbol{\theta}^*)$?
- **Solution**:

$$\begin{aligned}
p(y=1 \mid \mathbf{x}_0, \boldsymbol{\theta}^*) &= \sigma(\boldsymbol{\theta}^{*\mathrm{T}} \mathbf{x}_0) \\
&= \sigma(-0.5) \\
&= \frac{\exp(-0.5)}{1+\exp(-0.5)} \\
&\approx 0.378
\end{aligned}$$

Quiz: Logistic Regression

- Assume we have learned a binary logistic regression model $f:\mathbb{R}^2 \to [0,1]$ of the form

  $$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(\theta_1 x_1 + \theta_2 x_2)$$

  by maximizing the conditional log-likelihood on a data set

  $$\mathcal{D} = \{(\mathbf{x}_1, y_1),\ldots,(\mathbf{x}_N, y_N)\}$$

  as explained in the lecture. Assume the learned parameter vector is given by

  $$\boldsymbol{\theta}^* = (0.5, -1)$$

- **Question 2**: What can we say about the conditional likelihood of $\boldsymbol{\theta}^*$ and the model $\boldsymbol{\theta}^+ = 2\boldsymbol{\theta}^* = (1,-2)$?
  - It holds that $p(y_1,\ldots y_N \mid \mathbf{x}_1,\ldots,\mathbf{x}_N, \boldsymbol{\theta}^*) < p(y_1,\ldots y_N \mid \mathbf{x}_1,\ldots,\mathbf{x}_N, \boldsymbol{\theta}^+)$
  - It holds that $p(y_1,\ldots y_N \mid \mathbf{x}_1,\ldots,\mathbf{x}_N, \boldsymbol{\theta}^*) > p(y_1,\ldots y_N \mid \mathbf{x}_1,\ldots,\mathbf{x}_N, \boldsymbol{\theta}^+)$
  - It holds that $p(y_1,\ldots y_N \mid \mathbf{x}_1,\ldots,\mathbf{x}_N, \boldsymbol{\theta}^*) = p(y_1,\ldots y_N \mid \mathbf{x}_1,\ldots,\mathbf{x}_N, \boldsymbol{\theta}^+)$
  - We cannot draw any conclusions based on the information we have

Quiz: Logistic Regression

- Assume we have learned a binary logistic regression model $f:\mathbb{R}^2 \to [0,1]$ of the form

  $$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(\theta_1 x_1 + \theta_2 x_2)$$

  by maximizing the conditional log-likelihood on a data set

  $$\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$$

  as explained in the lecture. Assume the learned parameter vector is given by

  $$\boldsymbol{\theta}^* = (0.5, -1)$$

- **Solution**: It holds that $p(y_1,\ldots y_N \mid \mathbf{x}_1,\ldots,\mathbf{x}_N,\boldsymbol{\theta}^*) > p(y_1,\ldots y_N \mid \mathbf{x}_1,\ldots,\mathbf{x}_N,\boldsymbol{\theta}^+)$ because the model parameter vector $\boldsymbol{\theta}^*$ has been trained to maximize the conditional likelihood

Quiz: Logistic Regression

- Assume we have learned a binary logistic regression model $f:\mathbb{R}^2 \to [0,1]$ of the form

  $$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(\theta_1 x_1 + \theta_2 x_2)$$

  by maximizing the conditional log-likelihood on a data set

  $$\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$$

  as explained in the lecture. Assume the learned parameter vector is given by

  $$\boldsymbol{\theta}^* = (0.5, -1)$$

- **Question 3**: What can we say about the classification accuracy of the model $\boldsymbol{\theta}^*$ and the model $\boldsymbol{\theta}^+ = 2\boldsymbol{\theta}^* = (1, -2)$ on the training data?
  - Model $\boldsymbol{\theta}^+$ is more accurate than model $\boldsymbol{\theta}^*$
  - Model $\boldsymbol{\theta}^+$ is less accurate than model $\boldsymbol{\theta}^*$
  - The models $\boldsymbol{\theta}^+$ and $\boldsymbol{\theta}^*$ have exactly the same accuracy
  - We cannot draw any conclusions based on the information we have

Quiz: Logistic Regression

- Assume we have learned a binary logistic regression model $f:\mathbb{R}^2 \to [0,1]$ of the form

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(\theta_1 x_1 + \theta_2 x_2)$$

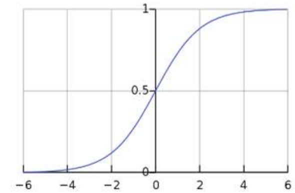
by maximizing the conditional log-likelihood on a data set

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$$

as explained in the lecture. Assume the learned parameter vector is given by

$$\boldsymbol{\theta}^* = (0.5, -1)$$

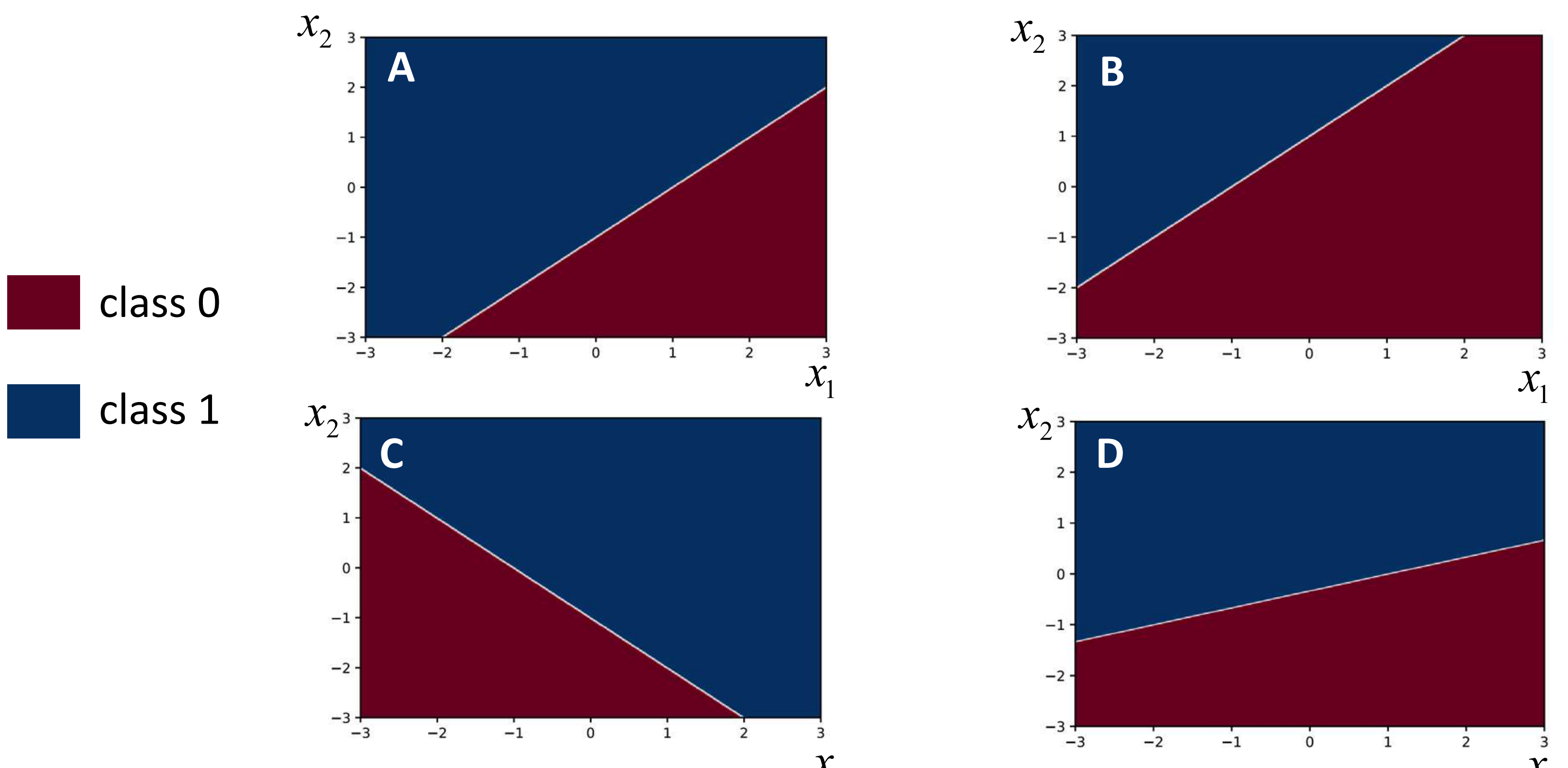
- **Question 3**: What can we say about the classification accuracy of the model $\boldsymbol{\theta}^*$ and the model $\boldsymbol{\theta}^+ = 2\boldsymbol{\theta}^* = (1,-2)$ on the training data?
- **Solution**:
  Both models have exactly the same classification accuracy. The reason is that both models make the same predictions, just with different confidence:
  - A logistic regression model predicts the positive class if $p(y=1 \mid \mathbf{x}, \boldsymbol{\theta}^*) \geq 0.5$, which is equivalent to $\mathbf{x}^{\mathrm{T}}\boldsymbol{\theta} \geq 0$
  - Of course, $\mathbf{x}^{\mathrm{T}} 2\boldsymbol{\theta} \geq 0 \Leftrightarrow \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta} \geq 0$



Quiz: Visualizing Linear Model

- Assume two dimensional data, $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$
- Which decision boundary matches the following binary logistic regression model:

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(-x_1 + x_2 + 1)$$



Solution: Visualizing Linear Model

- **Solution**:
- The points on the decision boundary are given by the condition

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(-x_1 + x_2 + 1) = 0.5$$

which implies

$$-x_1 + x_2 + 1 = 0.$$

- From this condition, we can derive the condition

$$x_2 = x_1 - 1.$$

- This is clearly the line in Figure A.

Solution: Visualizing Linear Model

- **Alternative Solution**:
- For the point $\mathbf{x} = \begin{pmatrix} -3 \\ -3 \end{pmatrix}$ the model predicts class 1:

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma(-(-3) + (-3) + 1) = \sigma(1) = \frac{e}{1+e} > 0.5$$

- Therefore, it has to be Figure A.