Foundations of Data Science:
Logistic regression -
The logistic regression classifier

Converting logistic regression to a classifier

$$\ln \frac{P(y=1 \mid \underline{x})}{1 - P(y=0 \mid \underline{x})} = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \dots$$

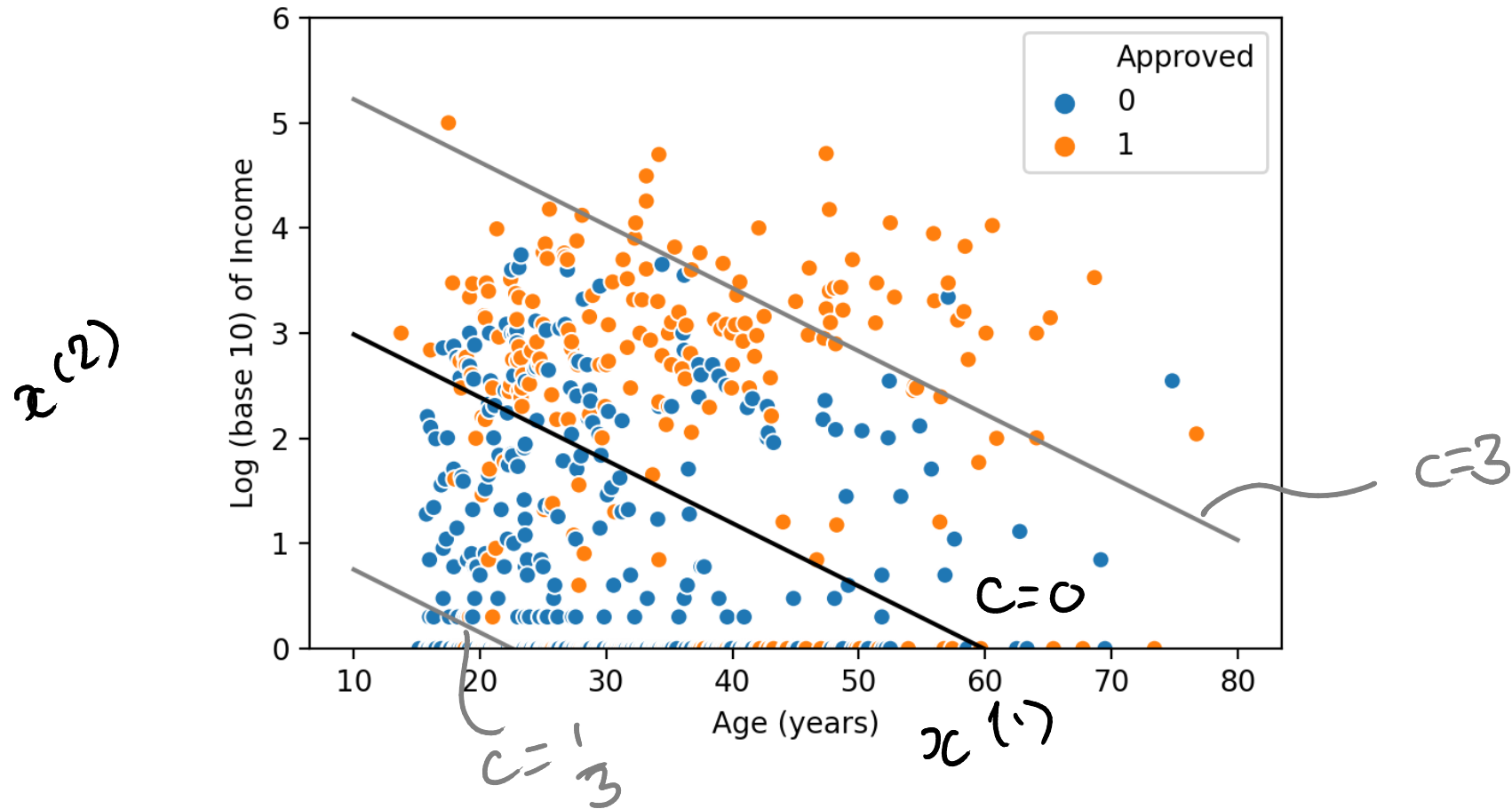$$\underline{x} = (x^{(1)}, x^{(2)}, \dots x^{(k)})$$

$$\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \dots \geq c \implies \hat{y} = 1$$

$$\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \dots < c \implies \hat{y} = 0$$

$$c = 0 \implies \text{odds of } 1$$

$$\implies p = 0.5$$

Decision boundary

$$\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots > c \Rightarrow \hat{y} = 1$$

$$\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots \leq c \Rightarrow \hat{y} = 0$$

Ethics: Logistic regression can be transparent

- If you are in employment you score 1.625, if not you score 0 $\hat{\beta}_1$

- Multiply your age by 0.029 and add the result to your score $\hat{\beta}_2$

- Round your income to the nearest 1000. Multiply the number of zeros in this figure by 0.320 and add the result to your score[5] $\hat{\beta}_3$

- If you scored more than 2.246, your credit will be approved $-\hat{\beta}_0$

Logistic regression versus k − NN