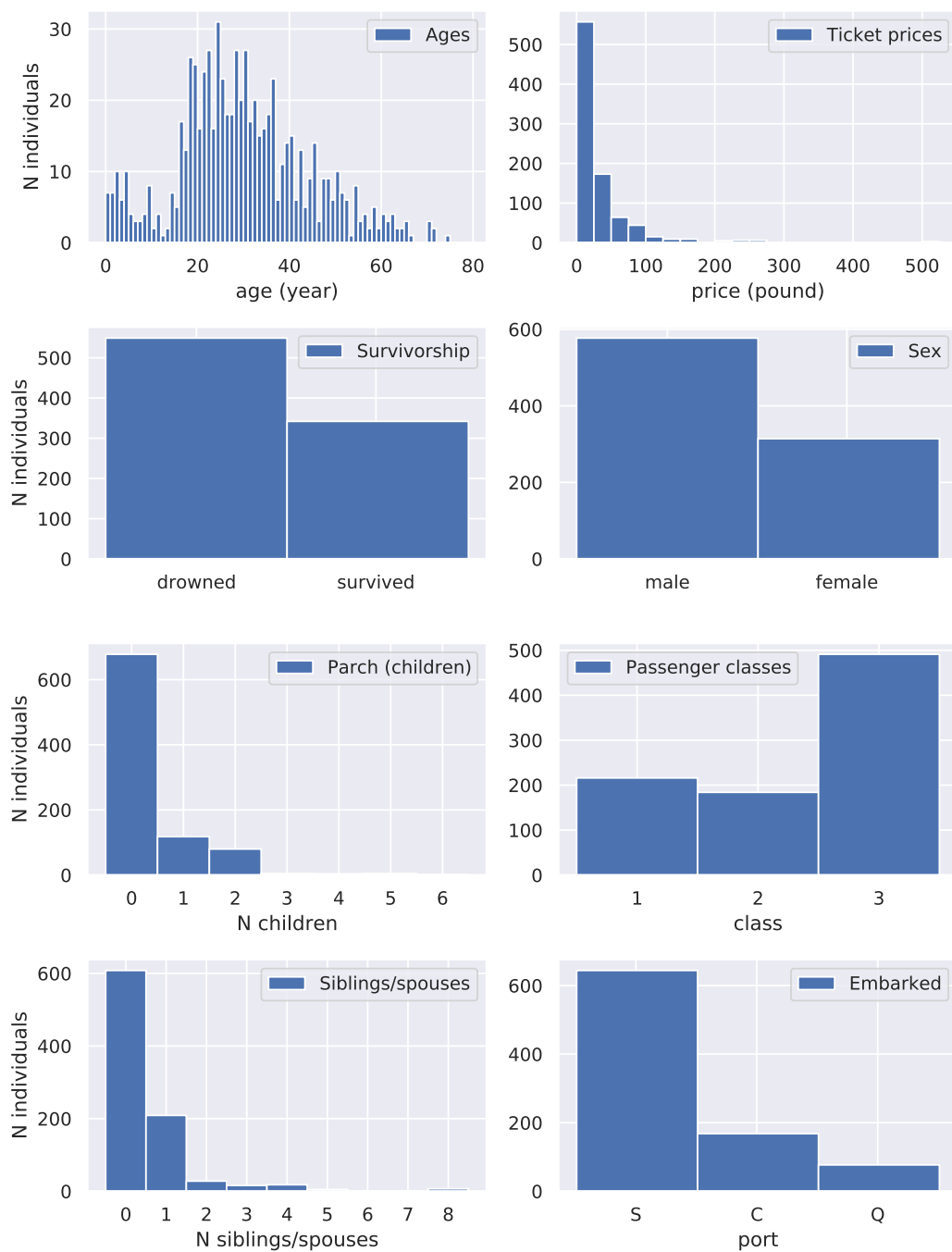


Analysis of Titanic data

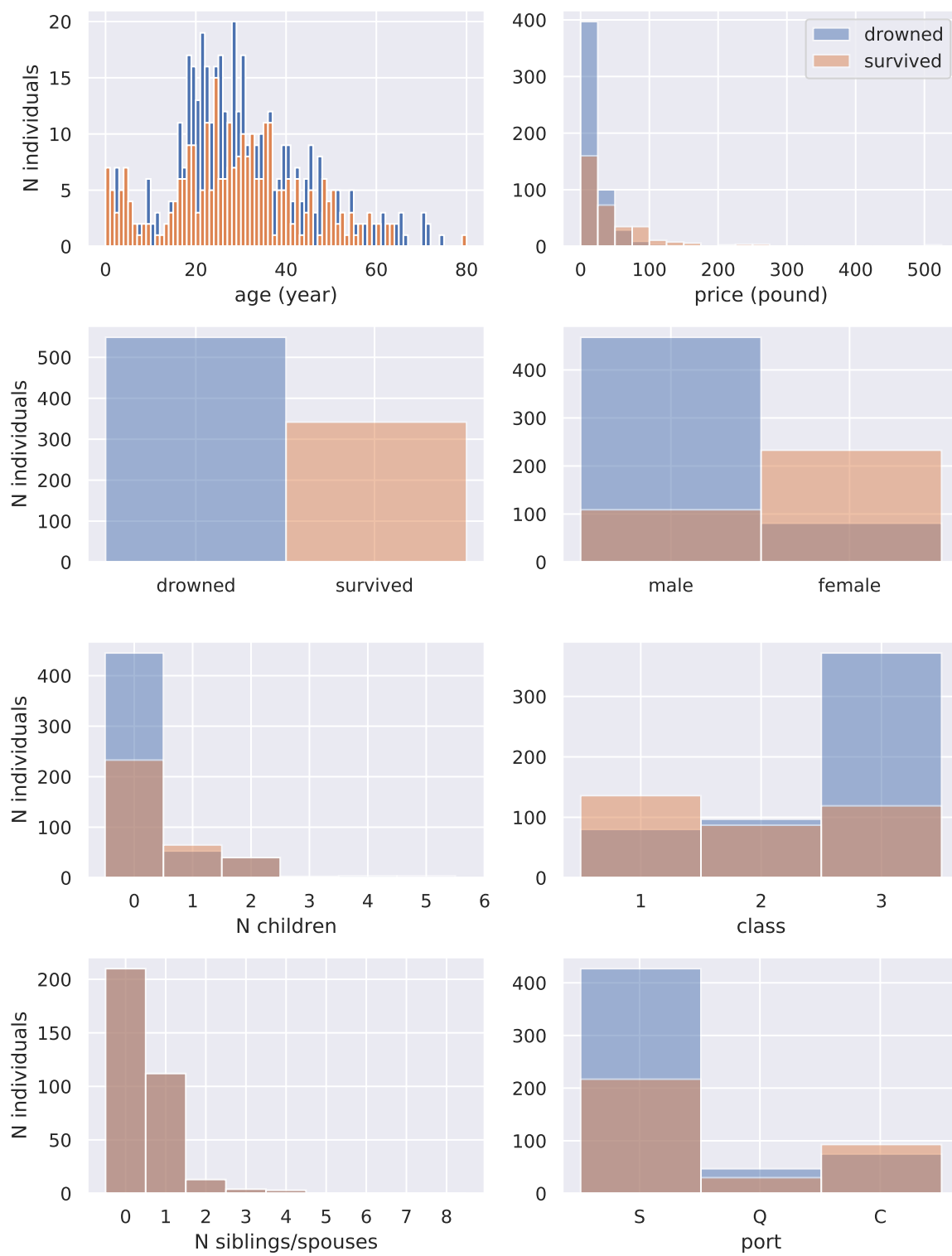
Samuel Knudsen

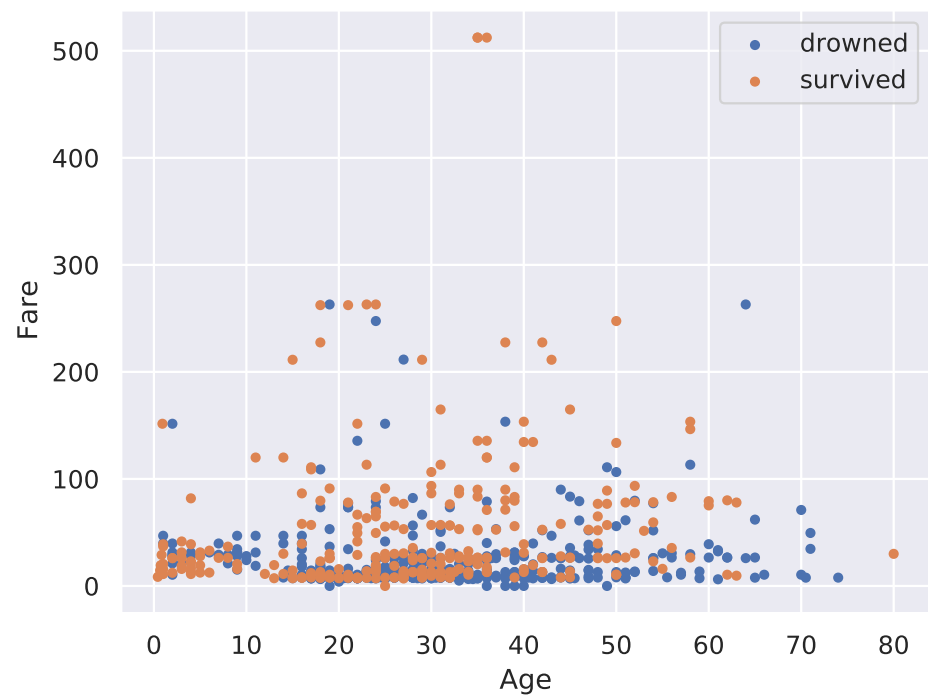
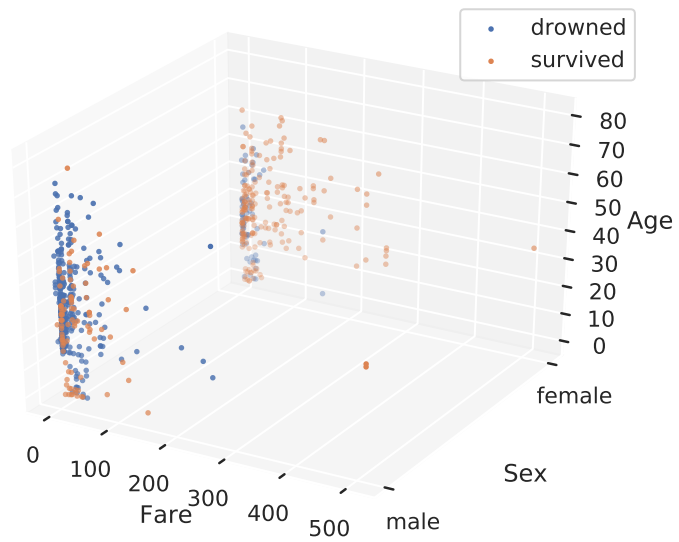
Last update: September 12, 2021

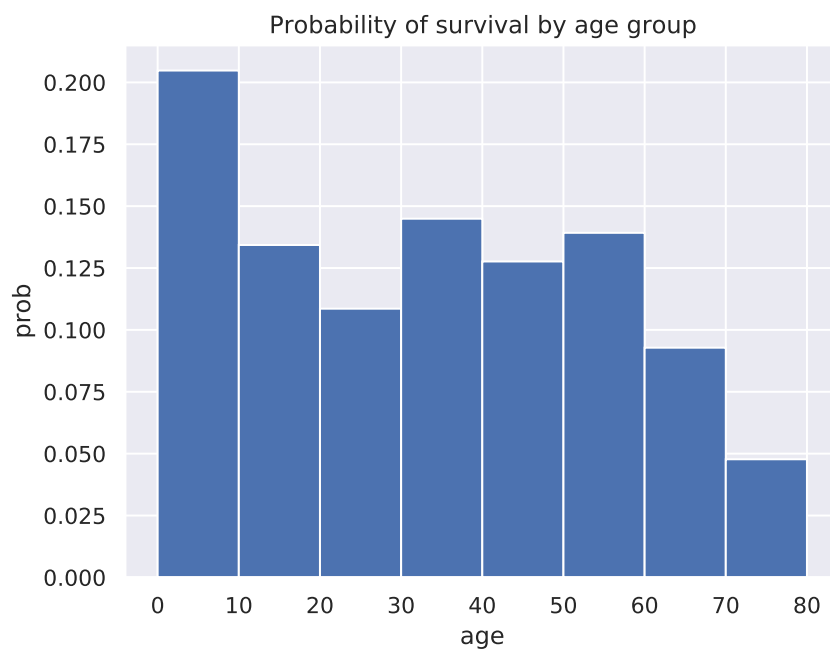
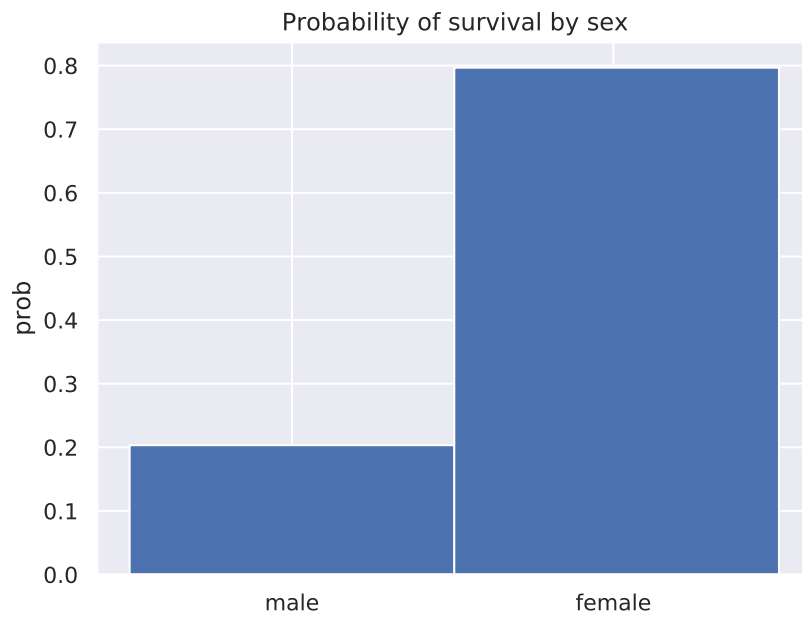
Parameter distributions

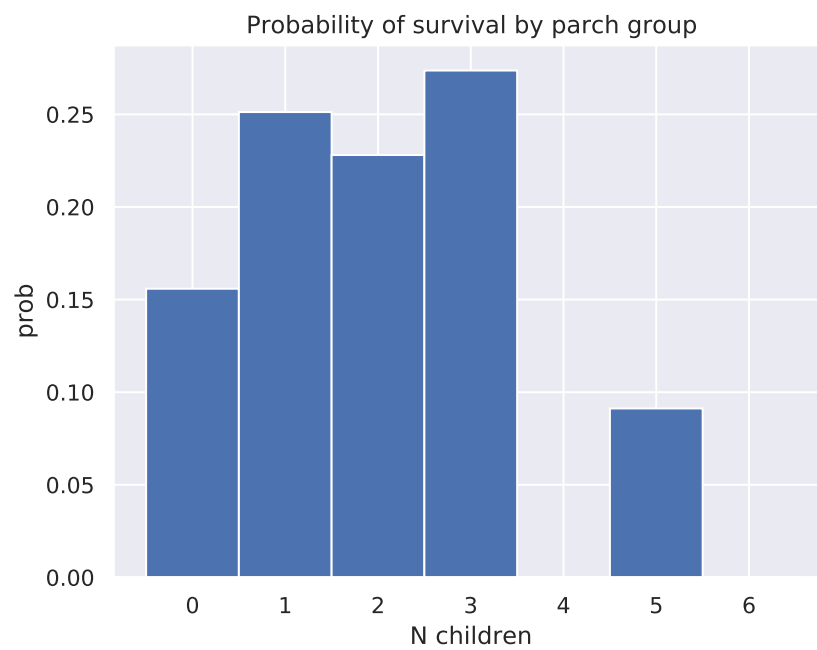
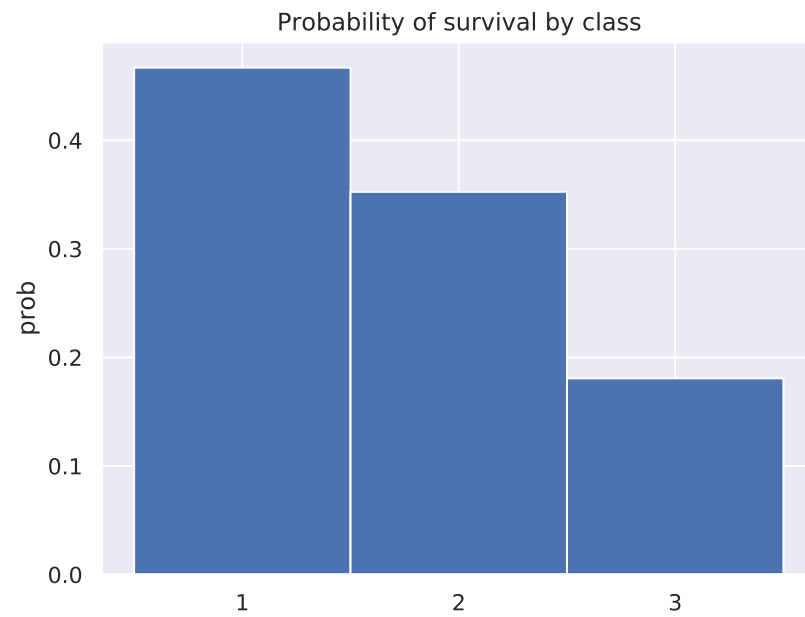


Parameter distributions by survivor status



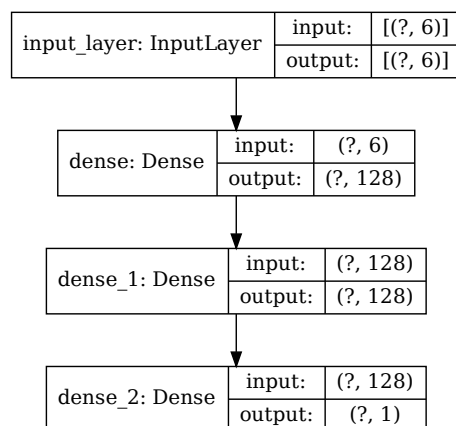






Training a neural network to predict survival

Need to represent strings as numbers. Zero-mean and normalize data.



Prediction just by sex

Wrote a predictor that predicts drowning if man, survival if woman. Gets 78% accuracy.

Bayesian probability analysis

Let us look at which piece of data about an individual gives us most information about the likelihood of survival, given our dataset D .

For some choice of $x = 0, 1$, $y = 0, 1$, we have by Bayes rule:

$$\text{Prob}(\text{survived} = x | \text{sex} = y, D) \propto \text{Prob}(\text{sex} = y | \text{survived} = x, D) \cdot \text{Prob}(\text{survived} = x | D) \quad (1)$$

We can calculate the prior (the probability estimate of survival prior to knowing the sex of the given person)

$$\text{Prob}(\text{survived} = x | D) = \frac{N_{\text{survived}=x}}{N_{\text{total passengers}}}, \quad (2)$$

and similarly the likelihood

$$\text{Prob}(\text{sex} = y | \text{survived} = x, D) = \frac{N_{\text{survived}=x, \text{sex}=y}}{N_{\text{total survivors}}} \quad (3)$$

with the normalization condition

$$\text{Prob}(\text{survived} = 0 | \text{sex} = y, D) + \text{Prob}(\text{survived} = 1 | \text{sex} = y, D) = 1. \quad (4)$$

$$\begin{aligned} p(A, B|D) &= p(A|B, D)p(B|D) \\ p(B, A|D) &= p(B|A, D)p(A|D) \end{aligned} \tag{5}$$

and

$$p(A, B|D) = p(B, A|D) \tag{6}$$

meaning

$$p(A|B, D) = \frac{p(B|A, D)p(A|D)}{p(B|D)} \tag{7}$$