Logistic regression unit:

- Derivation from linear regression (last week Friday)
- Lab activity: Applying logistic regression (Monday)
- Multiclass classification (**today**)

Today:

- The math of logistic regression
- Implementing logistic regression
- Multiclass classification

Logarithm rules:

$$\log xy = \log x + \log y \qquad \log x^c = c \log x$$

Differentiation rules:

$$\frac{d}{dx}\, c = 0 \qquad \frac{d}{dx}\, c\ f(x) = c\ f'(x)$$

$$\frac{d}{dx}\,(f(x)+g(x)) = f'(x)+g'(x) \qquad \frac{d}{dx}\, f \circ g(x) = (f' \circ g(x))\ g'(x)$$

$$\frac{d}{dx}\, f(x)g(x) = f(x)g'(x)+g(x)f'(x) \qquad \frac{d}{dx}\, \frac{f(x)}{g(x)} = \frac{g(x)f'(x)-f(x)g'(x)}{(g(x))^2}$$

$$\frac{d}{dx}\, \ln x = \frac{1}{x} \qquad \frac{d}{dx}\, \sigma(x) = (1-\sigma(x))(\sigma(x))$$

Model family for logistic regression (as a probability function):

$$p(\boldsymbol{x}) = \sigma(\boldsymbol{\theta}^T \boldsymbol{x})$$

Loss function (*mean log loss*):

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{n=0}^{N-1} \Big(y_n \ln \sigma(\boldsymbol{\theta}^T \boldsymbol{x_n}) + (1 - y_n) \ln(1 - \sigma(\boldsymbol{\theta}^T \boldsymbol{x_n})) \Big)$$

Gradient of the mean log loss (where i ranges over $[0, D]$, that is, as an index indicating a component to the gradient):

$$\nabla_{\boldsymbol{\theta}} \mathcal{L} = \frac{1}{N} \left[\cdots \sum_{n=0}^{N-1} x_{n,i} (\sigma(\boldsymbol{\theta}^T \boldsymbol{x_n}) - y_n) \cdots \right]$$

Coming up:

Due Wed, Feb 12:
Do gradient descent for linear regression programming assignment

Due Thurs, Feb 13:
Take logistic regression quiz

Due Wed, Feb 19:
Do logistic regression programming assignment

Due Thurs, Feb 20:
Read from Chapter 2 in the textbook, about probability
(See Canvas for details)

Due Fri, Feb 21:
Submit "Dataset" checkpoint for term project