Logistic regression unit:

- Derivation from linear regression (last week Friday)
- Lab activity: Applying logistic regression (Monday)
- Multiclass classification (**today**)
- [Begin Gaussian mixture models unit (Friday)]

Today:

- Correction/clarification of linear regression formulas
- Warm-up: The derivative of the logistic function
- Training logistic regression
- Multiclass classification

[Original:] The closed form solution for plain old linear regression is

$$\boldsymbol{\theta}^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}$$

For ridge regression, it is

$$\boldsymbol{\theta}^T = (\mathbf{X}^T\mathbf{X} + \mathbf{A})^{-1}\mathbf{X}^T\boldsymbol{y}$$

The *mean square error*:

$$\mathcal{L}_{MSE}(\boldsymbol{\theta}) = \frac{1}{N}||\boldsymbol{y}^T - \boldsymbol{\theta}^T\mathbf{X}^T||^2 = \frac{1}{N}||\boldsymbol{y} - \mathbf{X}\boldsymbol{\theta}||^2$$

The gradient of this loss function:

$$\nabla_{\boldsymbol{\theta}}\mathcal{L} = \frac{1}{N}(-2\boldsymbol{y}^T\mathbf{X} + 2\boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X})$$

For ridge:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \tfrac{1}{N}||\boldsymbol{y}^T - \boldsymbol{\theta}^T\mathbf{X}||^2 + \alpha||\theta||^2 \\ \nabla_{\boldsymbol{\theta}}\mathcal{L} &= \tfrac{1}{N}(-2\boldsymbol{y}^T\mathbf{X} + 2\boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X}) + 2\alpha\boldsymbol{\theta} \end{aligned}$$

For LASSO:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \tfrac{1}{N}||\boldsymbol{y}^T - \boldsymbol{\theta}^T\mathbf{X}||^2 + \alpha||\theta||^1 \\ \nabla_{\boldsymbol{\theta}}\mathcal{L} &= \tfrac{1}{N}(-2\boldsymbol{y}^T\mathbf{X} + 2\boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X}) + 2\alpha\,(\text{sign}(\theta_i)) \end{aligned}$$

[Fixed/clarified:] The closed form solution for plain old linear regression is

$$\boldsymbol{\theta}^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}$$

For ridge regression, it is

$$\boldsymbol{\theta}^T = (\mathbf{X}^T\mathbf{X} + \mathbf{A})^{-1}\mathbf{X}^T\boldsymbol{y}$$

The *mean square error*:

$$\mathcal{L}_{MSE}(\boldsymbol{\theta}) = \frac{1}{N}||\boldsymbol{y}^T - \boldsymbol{\theta}^T\mathbf{X}^T||_2^2 = \frac{1}{N}||\boldsymbol{y} - \mathbf{X}\boldsymbol{\theta}||_2^2$$

The gradient of this loss function:

$$\nabla_{\boldsymbol{\theta}}\mathcal{L} = \frac{1}{N}(-2\boldsymbol{y}^T\mathbf{X} + 2\boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X})$$

For ridge:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \tfrac{1}{N}||\boldsymbol{y}^T - \boldsymbol{\theta}^T\mathbf{X}||_2^2 + \alpha||\theta||_2^2 \\
\nabla_{\boldsymbol{\theta}}\mathcal{L} &= \tfrac{1}{N}(-2\boldsymbol{y}^T\mathbf{X} + 2\boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X}) + 2\alpha\boldsymbol{\theta}
\end{aligned}$$

For LASSO:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \tfrac{1}{N}||\boldsymbol{y}^T - \boldsymbol{\theta}^T\mathbf{X}||_2^2 + \alpha||\theta||_1 \\
\nabla_{\boldsymbol{\theta}}\mathcal{L} &= \tfrac{1}{N}(-2\boldsymbol{y}^T\mathbf{X} + 2\boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X}) + 1\alpha\,(\text{sign}(\theta_i))
\end{aligned}$$

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:

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

Score function for each class $k \in [0, K)$ (where K is the number of classes:

$$s_k(\boldsymbol{x}) = \boldsymbol{\theta}_{\boldsymbol{k}}^T \boldsymbol{x}$$

Softmax function $\sigma : \mathbb{R}^K \rightarrow [0, 1]^K$ (this is a different $\sigma \ldots$):

$$\sigma(\boldsymbol{z})_k = \frac{e^{z_k}}{\sum_{j=0}^{K-1} e^{z_j}} \text{ for } k \in [0, K)$$

Coming up:

Do gradient descent /linear regression assignment (due end-of-day Thurs, Feb 9)

Take logistic regression quiz (due classtime Fri, Feb 10)

Read textbook portions about probability and distributions (due end-of-day Mon, Feb 13)

Do logistic regression assignment (due end-of-day Wed, Feb 15)

(Reading and response assignment forthcoming. . .)

Project dataset confirmation *(due end-of-day Fri, Feb 17)*