Linear regression unit:

- Simple linear regression with ordinary least squares (last week Monday)
- Lab activity: Linear regression (last week Wednesday)
- Newton's method and gradient descent (last week Friday and this week Monday)
- Training linear regression using gradient descent (**today**)
- [Begin logistic regression (Friday)]

Today:

- Counting the costs
- Revisiting loss functions
- Applying gradient descent

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:

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

**Coming up:**

*Read textbook sections on linear regression (due end-of-day Mon, Jan 30)*
*Do linear regression assignment (due end-of-day Tues, Jan 31)*

*Take gradient descent quiz (due classtime Fri, Feb 3)*

**Project proposal** *(due end-of-day Fri, Feb 3)*