Announcement

- Unit 3 & 4 Test on Thursday, 11/9.
  - All Unit 3 & 4 related topics
  - Same rules as the previous two unit tests.
  - Practice tests available on Canvas
- Assignment 7 deadline postponed to 11/13.
  - First problem related to Unit 4
- Starting today, grading will be done only based on your Canvas submission.
  - Missing parts of your submission won’t be accepted after the “Available Until” Time.
  - No more mailing of answers for grading.
  - In-effect for both activities and assignments starting today.

Unit 3 & 4 Review

SVD and Least Squares Problems

- **Problem:** $\min_w ||d - Aw||^2$ where SVD of A is $A = U\Sigma V^T$
- **Solution:** $w^* = (A^T A)^{-1} A^T d = V\Sigma^{-1} U^T d$

Regularization via Ridge Regression

- **Problem:** $\min_w ||d - Aw||^2 + \lambda ||w||^2$
- **Solution:** $w^* = (A^T A + \lambda I)^{-1} A^T d = V(\Sigma^2 + \lambda I)^{-1} U^T d$

Regularization via Truncated SVD

- $w^* = V_r \Sigma_{\mathrm{r}}^{-1} U_r^T d$

Unit 3 & 4 Review

Low Rank Decomposition $A \approx TW^T$ T and W only r cols

- Clustering (K-Means algorithm):
  - T columns contains the r centroids
  - W columns encode the cluster membership of each sample

- **SVD low rank approximation** (SVD of A: $\mathrm{A} = \mathrm{U}\Sigma V^T$)
  - $T = U_r$ first r left singular vectors
  - $W = \Sigma_{\mathrm{r}} V_r^T$ first r right singular vectors scaled by singular values.
  - Optimal low rank approximation (in terms of MSE)

Unit 3 & 4 Review

- ***Eigen decomposition*** of B: $BE = \Lambda E$ or $Be_i = \lambda_i e_i$
  - If $B = AA^T$ Left SV of A $\Leftrightarrow$ Eigenvecs of B
  - If $B = A^T A$ Right SV of A $\Leftrightarrow$ Eigenvecs of B
  - Singular values of A $\left(\sigma_i^2\right) \Leftrightarrow$ Eigen values of B (λ_i)

- Algorithms:
  - ***Power Iteration***: Repeatedly applying a matrix **A** to a random vector **x** and normalizing $\boldsymbol{x} \leftarrow \frac{\boldsymbol{Ax}}{||\boldsymbol{Ax}||^2}$ converges to the first principal component of **A.**
  - ***Page Rank Algorithm:*** *Top Eigenvector of Graph Transition Matrix* represents the steady-state distribution over pages.
  - ***Low Rank Matrix Completion*** via *Iterative Singular Value Thresholding.*

Activity 16

Before:

loss function

Regularizer

$$\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||^2 + \lambda||\boldsymbol{w}||_2^2$$

(L2 gives solutions with small 2-norm)

closed form solution

$$\boldsymbol{w}^* = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

What if we want some other regularizer?

(maybe only a few non-zero entries)

$$\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||^2 + \lambda r(\boldsymbol{w})$$

Or different a loss function all together?

$$\min_{\boldsymbol{w}} f(\boldsymbol{w})$$

In general: no closed-form solution ➡ Numerical methods

Gradient Descent

Main idea: use the gradient to head downhill

$$\begin{aligned}&\text{goal: } \min_{\boldsymbol{w}} f(\boldsymbol{w})\\ &\text{for } k = 1\ldots\\ &\quad \boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - \tau\nabla f(\boldsymbol{w})\end{aligned}$$

step size

Gradient descent for least-squares:

$$\begin{aligned}&\text{goal: } \min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2\\ &\text{for } k = 1\ldots\\ &\quad \boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - \tau(2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} - 2\boldsymbol{X}^T\boldsymbol{y})\end{aligned}$$

Convergence guaranteed if:

$$0 < \tau < \frac{1}{||\boldsymbol{X}||_{op}^2}$$

Steepest direction is normal to the contour!