HW14 CS 189

1.
Who else did you work with on this homework? In case of course events, just describe the group. How did you work on this homework? Any comments about the homework?

I worked on this homework with Ehimare Okoyomon, Prashanth Ganeth, and Daniel Mockaitis. We worked by getting together throughout the week and communicating on facebook.

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up}

Nicholas Lorio, 26089160

4. When should you use k-svd and how can you determine the optimal k?

K-SVD is widely used in applications such as image processing, audio processing, biology, and document analysis to name a few. Logically, one of the best ways to determine k is by formulating a way to assign an empirical measure of the outcome from the SVD. We could test the results against a set of predetermined queries for which we know all the relevant documents in the set. For SVD representations of document files, a common measure to determine k has been to use synonym tests (TOEFL test for instance).

2 2.a. (3) $\min_{\pi, \mu} \sum_{k=1}^{K} \sum_{i \in \pi_k} \|x_i - \mu_k\|_2^2$ π_i cluster
μ_i is centroid

Since each sample x_i is clustered in exactly one partition thus we are minimizing K distinct problems of the form:

① $\min_{\mu \in \mathbb{R}^{k \times d}} \sum_{i \in \pi_k} \|x_i - \mu_k\|_2^2$

② $\min_{Z \in \mathbb{R}^{n \times k}, \mu \in \mathbb{R}^{k \times d}} \|X - \tilde{Z}\mu\|_F^2 = \min \sum_{i=1} \sum_{j=1} |(X - \tilde{Z}\mu)_{ij}|^2$
s.t. $\|z_i\|_0 = 1 \ \forall i$ & $z_i \in \{0,1\}^K$

$\tilde{Z} = \begin{bmatrix} \tilde{z}_1^T \\ \vdots \\ \tilde{z}_n^T \end{bmatrix}$ • by the constraints all but one of these will be zero vector

$\underbrace{\tilde{Z}_c}_{\text{under constraints}} = \begin{bmatrix} \tilde{z}_i^T \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times k}$, $\underset{\text{n\times d}}{\tilde{Z}_c \mu} = \begin{bmatrix} 0 \\ \tilde{z}_i^T \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \mu_1^T \\ \vdots \\ \mu_k^T \end{bmatrix} = \tilde{z}_i^T \cdot \mu_{k \in K}$
$= 1 \cdot \mu_{k \in K}$

• By def^n of Frobenius norm
②
$= \min_{Z, \mu} \sum_i \sum_j |(X - \tilde{Z}\mu)_{ij}|^2 = \sum_i \sum |(X - \mu_{k \in K})_{ij}|$

$$= \min_{z,\mu} \sum \sum \left\| (x - \mu_{k \in k}) \right|^2$$

$$= \min \sum_{i \in \Pi k} \| x_i - \mu_k \|_2^2 \qquad = 1$$

189

2.a. (3) $\min_{\pi, \mu \in \mathbb{R}^{k \times 2}} \sum_{k=1}^{K} \sum_{i \in \pi_k} \| x_i - \mu_k \|_2^2$ $\left\{ \begin{array}{l} \pi_i \text{ is cluster} \\ \mu_i \text{ is the centroid} \end{array} \right.$

$\mu = \begin{bmatrix} \mu_1^T \\ \vdots \\ \mu_k^T \end{bmatrix}$ & $\tilde{Z} = \begin{bmatrix} \tilde{z}_1^T \\ \vdots \\ \tilde{z}_n^T \end{bmatrix}$ "Coefficient Matrix" $\mathcal{X} = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}$

- if we know $\pi_1 \ldots \pi_k$, the cluster assignments

=> we only need to determine the centroid locations

- the choice of centroid locations μ_i does not affect the distances of pts in π_j to c_j for $i \neq j$, ∴ we can consider each cluster separately.
- each sample x_i is contained in exactly one partition.

$\min_{\mu \in \mathbb{R}^{k \times 2}} \sum_{i \in \pi_k} \| x_i - \mu_k \|_2^2$

- $(x_i - \mu_k)^T (x_i - \mu_k) = (x_i^T x_i + \mu_k^T \mu_k - 2 \mu_k^T x_i^T)$

Using eqn 2) $\propto \sum_{i \in \pi_k} \left(\sum_{k=1}^{K} z_{ik} \mu_k^T x_i + \mu_k^T \mu_k - 2 \mu_k^T \sum_{k=1}^{K} z_{ik} \mu_k^T \right)$

2.b. for $j = 1, 2, \dots, S$ for algo 3

constant of D_k

Show $\min_{\beta \in \mathbb{R}} , \min_{k \in [K]} \| X - \sum_{D_k \in B_{j-1} \cup \{D_k\}} \beta_k D_k \|_2^2$

is non increasing i.e. can do as OMP?

gradient is not positive?

* assuming fixed dictionary D, each sparse step in the algorithm decreases total error $\|X - \tilde{Z}D\|_F^2$. @ the update step for d_k an addition reduction or no change is guarantee for MSE that does not violate sparsity constraint. This ensures that a series of these steps will have monotone MSE reduction
-> $\therefore$ converge to a local min is guaranteed.

* we can ensure convergence by external interference

we are saying the best solution we got whats given in the new iteration. we update to better one. This way we always get an improvement.

* k-SVD: An algorithm for designing overcomplete dictionary for sparse representation
Michal Aharon, Michael Elad, Alfred Bruckstein

2.c.

∘From (2b) we know that the min value of the obj func is non increasing so ∴ we can see

z_i for (D, x_i, S) has an z_ℓ as the optimal coeffd of D_ℓ.

-> we then compare the new iteration outcome z_i' w/ the previous iterations outcome which means

$$\|x_i^T - z_i^T D\|_2 > \|x_i^T - (z_i')^T D\|_2$$

then z is updated to z_i' (if ↑ true)

∴ if no z_i' makes this inequality true then we are left w/ the same outcome for the obj func $\|x - zD\|_F^2$.

If my z_i' completes the inequality then we have a new z' w/ different z values since this z' ensures $\|x_i^T - (z_i')^T D\|_2$ is small than the og. z_i's value we can't increase the overall objective function

∴ Algo 2 can't increase the value of obj func 1. They will either be the same or decrease.

3.

5) $\ell(w) = \|y - Xw\|_2^2 = \sum_{i=1}^{n} (y_i - x_i^T w)^2$

6) $w_{t+1} = w_t - \alpha \nabla_w \left[\left(y_i - \frac{1}{p} (x_i \odot M)^T w \right)^2 \right]$ $M \sim \text{Bernoulli}(p)$ i.e.

a. $X = [x^{(1)} \ x^{(2)}]$ $p = 2/3$ Find X', Y'
s.t. Same as dropout on
o.g. Data.

$$X' = \begin{bmatrix} \alpha_1 x^{(1)} & \alpha_1 x^{(2)} \\ \alpha_2 x^{(1)} & 0 \\ 0 & \alpha_3 x^{(2)} \\ 0 & 0 \end{bmatrix} \qquad Y' = \begin{bmatrix} \alpha_1 Y \\ \alpha_2 Y \\ \alpha_3 Y \\ \alpha_4 Y \end{bmatrix}$$

Show Prob(Mag each update is Same for augment & norm data)

.(b)

w'

X, Y

$$w_{t+1} = w_t - \alpha \nabla_w \left[(y_i - x_i^T w)^2 \right]$$

X', Y'

$$\nabla_w = 2 (y_i - x_i^T w)' (-x_i)$$
$$= -2 (y_i - x_i^T w) x_i$$

$$w_{t+1} = w_t + 2\alpha (y_i - x_i^T w_t) x_i$$

$i \sim U(1, n)$