Recitation (Review Week)

Alex Dong

CDS, NYU

Fall 2020

Review day: Linear Regression and SVD

- Your Questions
- Linear Regression Review
- SVD Review

Question: Linear Regression Review 1

- Goal is to minimize $\ell = \|X\beta - y\|_2^2$ w.r.t β
  - ℓ is loss
  - X is your data/design matrix
  - β is the coefficients that transform X into y

Let $x_1, .., x_n \in \mathbb{R}^d$, $y_1, .., y_n, \in \mathbb{R}$. Let $\beta \in \mathbb{R}^d$

1. Recall that X is defined to have $x_1, .., x_n$ as the rows of X, and y is defined to have $y_1, ..., y_n$ as its entries.
   Show that $\|X\beta - y\|_2^2 = \sum_{i=1}^n (y_i - \langle x_i, \beta \rangle)^2$
2. Notice that when $x_i = \vec{0}$, $\langle x_i, \beta \rangle = 0$. Suppose we wanted to add an 'intercept', or 'bias' $\beta_0 \in \mathbb{R}$, so that we are trying to minimize $\|X\beta - y\|_2^2 = \sum_{i=1}^n (y_i - \langle x_i, \beta \rangle - \beta_0)^2$.
   How can we modify our definitions of X, y, β to easily do this?

Solutions: Linear Regression Review 1

Solution

1. *Recall that X is defined to have $x_1, .., x_n$ as the rows of X, and y is defined to have $y_1, ..., y_n$ as its entries.*
   Show that $\|X\beta - y\|_2^2 = \sum_{i=1}^{n}(y_i - \langle x_i, \beta\rangle)^2$
   Use Inner product method of matrix multiplication
2. *Notice that when $x_i = \vec{0}$, $\langle x_i, \beta\rangle = 0$. Suppose we wanted to add an 'intercept', or 'bias' $\beta_0 \in \mathbb{R}$, so that we are trying to minimize $\|X\beta - y\|_2^2 = \sum_{i=1}^{n}(y_i - \langle x_i, \beta\rangle - \beta_0)^2$.*
   How can we modify our definitions of X, y, β to easily do this?
   Append/prepend a column of 1's to X, and append/prepend β_0.

Questions: Ridge Regression and Multicollinearity

Let $X \in \mathbb{R}^{n \times d}$, $n > d$, and *not have full rank.* (X is a data matrix)
Recall that the OLS solution is $\hat{x} = (X^T X)^{-1} X^T y$.

1. Since X is not full rank, what does this say about the features?
2. What is the issue with the OLS solution?
3. The ridge regression solution is given by $(X^T X + \lambda Id_d)^{-1} X^T y$. How does this fix the issue?
4. Suppose that X has SVD $X = U \Sigma V^T$, and X has singular values $\sigma_1, ..., \sigma_d$. What are the eigenvalues of $X^T X + \lambda Id_d$?
5. How does increasing λ affect the condition number of $(X^T X + \lambda Id_d)$?

Solutions: Ridge Regression and Multicollinearity

Let $X \in \mathbb{R}^{n \times d}$, $n > d$, and *not have full rank.* (X is a data matrix)
Recall that the OLS solution is $\hat{x} = (X^T X)^{-1} X^T y$.

Solution

1. *Since X is not full rank, what does this say about the features?*
   Columns of X are not linearly independent, so some of the features can be perfectly explained by other features.
2. *What is the issue with the OLS solution?*
   Since X does not have full rank, $X^T X$ doesn't have full rank and is not invertible. So the OLS solution is not well-defined.
3. *The ridge regression solution is given by $(X^T X + \lambda Id_d)^{-1} X^T y$. How does this fix the issue?*
   Adding λId_d to $X^T X$ shifts its eigenvalues up, which makes $(X^T X + \lambda Id_d)$ invertible.

Solutions: Ridge Regression and Multicollinearity

Let $X \in \mathbb{R}^{n \times d}$, $n > d$, and *not have full rank.* (X is a data matrix)
Recall that the OLS solution is $\hat{x} = (X^T X)^{-1} X^T y$.

Solution

4. *Suppose that X has SVD $X = U\Sigma V^T$, and X has singular values $\sigma_1, ..., \sigma_d$. What are the eigenvalues of $X^T X + \lambda Id_d$?*
   Note that $X^T X = V\Sigma^T \Sigma V^T$
   Eigvals of $X^T X$: $\sigma_1^2, ..., \sigma_d^2$, (Note: X isn't full rank, so $\sigma_d = 0$)
   Eigvals of $X^T X + \lambda Id_d$: $\sigma_1^2 + \lambda, ..., \sigma_d^2 + \lambda$.
5. *How does increasing λ affect the condition number of $(X^T X + \lambda Id_d)$ vs $X^T X$?*
   Condition number of $X^T X = \frac{\sigma_1^2}{\sigma_d^2} = \infty$
   Condition number of $(X^T X + \lambda Id_d) = \frac{\sigma_1^2 + \lambda}{\sigma_d^2 + \lambda}$
   Furthermore, for $\lambda_1 > \lambda_2$, we get the relationship
   $\frac{\sigma_1^2 + \lambda_1}{\sigma_d^2 + \lambda_1} < \frac{\sigma_1^2 + \lambda_2}{\sigma_d^2 + \lambda_2}$

Questions: More SVD

Let $A \in \mathbb{R}^{n \times m}$, where $n > m$, have SVD $A = U\Sigma V^T$.
Let $e_{i,m}, e_{i,n}$ denote the ith standard basis vector in $\mathbb{R}^m$, and $\mathbb{R}^n$.

1. Give basis transformations for U, Σ, V^T.
   (Your answer should look like $T(a) = b$. where you select convenient a's that form a basis for the origin space of T. Be careful about the dimensions w/ Σ!)

2. Let $x \in \mathbb{R}^m$ s.t $x = \sum_{i=1}^{m} \alpha_i v_i$, where v_i are the columns of V. and write the expressions for $V^T x$, $\Sigma V^T x$, and $U\Sigma V^T x$.

3. Use this to show $Im(AA^T) = Im(A)$

Solutions 1: More SVD

Let $A \in \mathbb{R}^{n \times m}$, where $n > m$, have SVD $A = U\Sigma V^T$.
Let $e_{i,m}, e_{i,n}$ denote the ith standard basis vector in $\mathbb{R}^m$, and $\mathbb{R}^n$.

1. Give basis transformations for U, Σ, V^T.
   (Your answer should look like $T(a) = b$. where you select convenient a's that form a basis for the origin space of T. Be careful about the dimensions w/ Σ!)

Solution

$V^T(v_i) = e_{i,m} \quad \textit{for } i \in \{1, ..., m\}$
$\Sigma(e_{i,m}) = \sigma_i e_{i,n} \quad \textit{for } i \in \{1, ..., m\}$ *(Note the m here)*
$U(e_{i,n}) = u_i \quad \textit{for } i \in \{1, ..., n\}$

Solutions 2: More SVD

Let $A \in \mathbb{R}^{n \times m}$, where $n > m$, have SVD $A = U\Sigma V^T$.
Let $e_{i,m}, e_{i,n}$ denote the ith standard basis vector in $\mathbb{R}^m$, and $\mathbb{R}^n$.

2. Let $x \in \mathbb{R}^m$ s.t $x = \sum_{i=1}^{m} \alpha_i v_i$, where v_i are the columns of V. Write the expressions for $V^T x$ $\Sigma V^T x$, and $U\Sigma V^T x$.

Solution

Let $e_{i,m}$ denote the ith standard basis vector in $\mathbb{R}^m$.

$x = \sum_{i=1}^{m} \alpha_i v_i$

$V^T x = \sum_{i=1}^{m} \alpha_i e_{i,m}$

$\Sigma V^T x = \sum_{i=1}^{min(m,n)} \sigma_i \alpha_i e_{i,n}$

$U\Sigma V^T x = \sum_{i=1}^{min(m,n)} \sigma_i \alpha_i u_i$

Solutions 3: More SVD

Let $A \in \mathbb{R}^{n \times m}$, where $n > m$, have SVD $A = U\Sigma V^T$.
Let $e_{i,m}, e_{i,n}$ denote the ith standard basis vector in $\mathbb{R}^m$, and $\mathbb{R}^n$.

3. Use this to show $Im(AA^T) = Im(A)$

Solution

Let $x \in \mathbb{R}^m, x = \sum_{i=1}^m \alpha_i v_i$. *Then* $Ax = \sum_{i=1}^m \alpha_i \sigma_i u_i$.
Let $y \in \mathbb{R}^n, y = \sum_{i=1}^n \beta_i u_i$. *Then* $AA^T y = U\Sigma\Sigma^T U^T y = \sum_{i=1}^m \beta_i \sigma_i^2 u_i$.
We show $Im(A) \subset Im(AA^T)$
Now, let $p \in \mathrm{Im}(A) \subset \mathbb{R}^n$.
z *must be of the form* $p = \sum_{i=1}^m \gamma_i u_i$.
Notice that for $q \in \mathbb{R}^m$, *where*

$$q = \sum_{i=1}^m c_i v_i, \quad \textit{and} \quad c_i = \frac{\gamma_i}{\sigma_i} \textit{ if } \gamma_i \neq 0,\ c_i = 0 \textit{ if } \gamma_i = 0$$

We get $Aq = p$. *Similarly, if we let* $r \in \mathbb{R}^n$ *where*

$$r = \sum_{i=1}^m d_i v_i, \quad \textit{and} \quad d_i = \frac{\gamma_i}{\sigma_i^2} \textit{ if } \gamma_i \neq 0,\ d_i = 0 \textit{ if } \gamma_i = 0$$

We get $AA^T r = p$.
So, we have found a vector r, *s.t* $AA^T r = p \in Im(A)$. *Then* $Im(A) \subset Im(A^T)$.