Lab 10

DSGA-1014: Linear Algebra and Optimization

CDS at NYU
Zahra Kadkhodaie

Fall 2021

A with linearly independent columns

Assume $A \in R^{n \times n}$ is full rank. We have learned that $Ax = b$ has a unique solution. Describe the solution from a column and a row perspective.

A with linearly independent columns

Assume $A \in R^{n \times m}$ is full rank, where $n > m$ (i.e. A is a tall matrix). In this case $Ax = b$ is a system of equations with too many rows (i.e. more equations than variables). We call this system of equations over-determined, which happens a lot in practice. Describe the solution of this system.

A with linearly independent columns

How do we solve $Ax = b$ when $b \notin Im(A)$? We can't! So we compromise and find the next best vector: x_{LS} such that Euclidean distance between Ax_{LS} and b is minimum. That is, x_{LS} is mapped to a vector on $Im(A)$ which is as close to b as possible.

$$x_{LS} = \text{argmin}_x ||Ax - b||^2$$

Show that Ax_{LS} is the projection of b onto $Im(A)$.
Show that error, $e = Ax - b$, is in $Im(A)^{\perp}$.

A with linearly dependent columns

The least square solution presented above only works if $A^T A$ is invertible. Since the $Ker(A^T A) = Ker(A)$, the least square as defined above only exists when columns of A are independent. Describe possible solutions.

A with linearly dependent columns: pseudo-inverse

To solve this problem, we define pseudo-inverse as $A^\dagger = V\Sigma' U^T$ where $\Sigma' \in R^{d\times n}$ with $\Sigma'_{ii} = 1/\Sigma_{ii}$ if $\Sigma_{ii} \neq 0$, and zero otherwise. Show that $A^\dagger \in R^{d\times n}$ is the only matrix in $R^{d\times n}$such that

1. $AA^\dagger A = A$
2. $A^\dagger A A^\dagger = A^\dagger$
3. $AA^\dagger \in R^{n\times n}$ and $A^\dagger A \in R^{d\times d}$ are symmetric matrices.

A with linearly dependent columns: pseudo-inverse

Using pseudo-inverse, we define the least square solution as $x_{LS} = A^{\dagger} y$.

1. Show that when columns of A are independent the two least square solutions are the same.
2. Show that x_{LS} is always in the row space of A.
3. Give the set of all vectors that minimize $||Ax - y||^2$?

A with linearly dependent columns: pseudo-inverse

Note: pseudo-inverse is particularly useful when $A \in R^{n \times m}$ is a short matrix ($n < m$). In this case, $Ax = b$ is an under-determined system of equations and even if A is full rank, rank(A) = n, columns of A are not independent and $A^T A$ is not invertible.

Ridge regression

Sometimes the objective deviates from least square solution. In Ridge regression, we add a penalty term to least square objective to promote a solution with small norm.

$$x_{ridge} = \text{arg min }_x ||Ax - b||^2 + \lambda ||x||^2$$

Show that x_{ridge} is in the row space of A.

Ridge regression

Ax_{ridge} is no longer an orthogonal projection of b onto the Im(A). It is a modified projection where the component of the data in the direction of each left singular vector of the feature matrix is shrunk by a factor of $\sigma_i^2/(\sigma_i^2+\lambda)$ where σ_i is the corresponding singular value. Show that

$$Ax_{ridge} = \sum_{i=1}^{m} \frac{\sigma_i^2}{\sigma_i^2+\lambda} \langle b, u_i \rangle u_i$$

where u_i are the left singular vectors of A.