# Least Square

## Least Squares Problem

- ▼ Keywords
    - ▼ Inner product (= dot product)
    - ▼ Vector norm
    - ▼ Unit vector
    - ▼ Orthogonal vectors

### Over-determined Linear Systems

1. Over-determined linear systems
    a. the number of equations >> the number of variables
    b. usually no solution exists
        i. b/c b vector is not included in the span of column vectors of the matrix A
2. Motivation for least squares
    a. approximately obtain the solution
    b. minimize the sum of squared errors (= least squares)
        i. error = $b - Ax$
        ii. the sum of squared errors = $||b - Ax||$
3. Least squares problem
    a. given an over-determined system $Ax = b$
    b. obtain least squares solution $\hat{x}$
        i. $\hat{x} = argmin_x ||b - Ax||$

### Inner Product

1. Inner product (= dot product)
    a. element-wise multiplication + summation
    b. view as matrix multiplication

$$u * v = u^T v$$

- $*$ is denoted as dot for convenience

c. linear combination

$$(c_1 u_1 + ... + c_p u_p) * w = c_1(u_1 * w) + ... + c_p(u_p * w)$$

## Vector Norm

1. Vector norm (= vector length)
   a. non-negative scalar $||v||$ defined as the square root of $v * v$

   $$||v||^2 = v * v$$

   - This trick is widely used in machine learning

## Unit Vector

1. Unit vector
   a. a vector whose length is 1
   b. vector normalization
      i. make the length of a vector to 1

      $$u = (1/||v||) * v$$

## Distance between Vectors

1. Distance between $u$ and $v$
   a. written as $dist(u, v)$
   b. the length of the vector $u - v$

   $$dist(u, v) = ||u - v||$$

## Inner Product and Angle Between Vectors

1. Rewriting the inner product using norms and angle
   a. $u * v = ||u||||v||cos(\theta)$

## Orthogonal Vectors

1. Orthogonal (=perpendicular)
   a. the angle between two vectors is 0

$$u * v = ||u||||v||cos(\theta) = 0$$

- vectors $u, v$ are nonzero

# Least Squares & Geometrical Interpretation

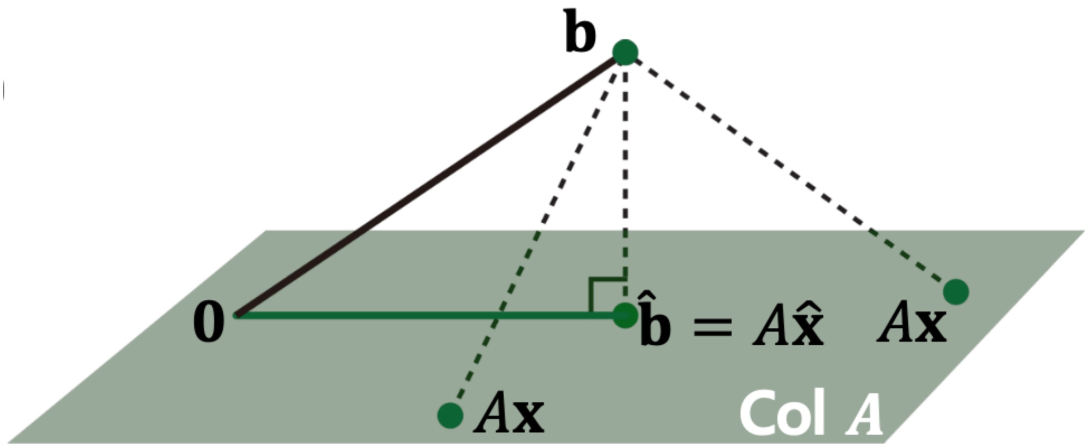
▼ Keywords

  ▼ Over-determined system

  ▼ Least squares

## Geometrical Interpretation of Least Squares

1. Geometrical interpretation
    a. find the shortest distance between $b, \hat{b}$
    b. the solution is orthogonal vector




        i. $b, \hat{b}$ should be orthogonal
        ii. $\hat{b}, Ax$ should be orthogonal
    c. how can we make the infinite number of solution to finite?
        a. use the basis
            i. $(b - \hat{b}) * (a_1x_1 + a_2x_2 + a_3x_3) = 0$
            ii. $A^T(b - A\hat{x}) = 0$

### Normal Equation

1. Normal equation
    a. given a least squares problem $Ax = b$
    b. obtain a normal equation
    $A^T A\hat{x} = A^T b$
        - which is equal to $A^T(b - A\hat{x})$
2. To be continue...

# Normal Equation

▼ Keywords

▼ Normal equation

### Normal Equation

1. Normal equation
    a. the normal equation can be viewed as a new linear system
    $Cx = d$
        - $C = A^T A, d = A^T b$
    b. if $C$ is invertible, the solution is computed as
    $\hat{x} = (A^T A)^{-1} A^T b$
    c. if $C$ is non-invertible, the solution is infinite
        i. normal equation is a special case, so there is no case for none solution
        ii. however, $C = A^T A$ is usually invertible
2. Another derivation of normal equation
    a. $\hat{x} = argmin_x ||b - Ax|| = argmin_x ||b - Ax||^2$
        i. $argmin_x (b - Ax)^T (b - Ax) = b^T b - x^T A^T b - b^T Ax + x^T A^T Ax$
    b. computing derivatives w.r.t $x$
        i. $-A^T b - A^T b + 2A^T Ax = 0$
    c. vector derivations commonly used at fc layer

i. $f(x) = a^T x = x^T a, f'(x) = a$

ii. $f * g = f' * g + f * g'$

# Orthogonal Projection

- Keywords
    - Orthogonality & Orthonormality
    - Orthogonal basis & Orthonormal basis
    - Orthogonal projection

## Orthogonal Projection Perspective

1. Orthogonal projection
    a. orthogonal projection of $b$ onto Col A
    $\hat{b} = f(b) = A\hat{x} = A(A^T A)^{-1} A^T b$
    b. when $A = U$, $A^T A = I$
        i. where $U$ consists of orthogonal vectors
        ii. $\hat{b} = UU^T b$

## Orthogonal and Orthonormal Sets

1. Orthogonal sets
    a. each pair of distinct vectors from the set is orthogonal
    b. 서로와 서로 간에 항상 수직을 이루는 벡터들의 집합
    c. is it always linearly independent?
        i. Yes
2. Orthonormal sets
    a. it is an orthogonal set of unit vectors
    b. orthogonal vector set 에서 주어진 방향은 그대로 두고 크기만 1인 벡터들의 집합

## Orthogonal and Orthonormal Basis

1. Orthogonal basis
    a. make basis $\{v_1, ..., v_p\}$ of a p-dimensional subspace W as an orthogonal

2. Orthonormal basis
    a. make the above basis as an orthonormal basis
3. Gram-Schmidt process (QR factorization)
    a. this process makes the basis as an orthogonal or orthonormal
    b. compute the orthogonal projection y onto W
    c. QR factorization
        i. matrix 를 분해하는 과정을 의미
    d. To be continue...
4. Orthogonal projection $\hat{y}$ of $y$ onto line
    a. 정사영된 벡터 $\hat{y}$ 의 길이

       $||y||cos\theta = y * u/||u||$
    b. $\hat{y} = proj_L y = (y * u/||u||^2)u = (y * u/u * u)u$
        i. if $u$ is a unit vector, $\hat{y} = proj_L y = (y * u)u$
            - $L$ means a line which is a span
5. Orthogonal projection $\hat{y}$ of $y$ onto plane
    a. projection is done independently on each orthogonal basis vector
    b. $\hat{y} = proj_L y = (y * u_1/u_1 * u_1)u_1 + (y * u_2/u_2 * u_2)u_2$
        i. if $u_1, u_2$ are unit vectors, $\hat{y} = proj_L y = (y * u_1)u_1 + (y * u_2)u_2$
    c. 삼수선의 정리
        i. 3개의 선에 대하여 2개의 선이 수직이면 나머지 1개도 수직이다

## Transformation: Orthogonal Projection

1. Transformation of orthogonal projection $\hat{b}$ of $b$
    a. given orthonormal basis $u_1, u_2$
    b. $\hat{b} = f(b) = (b * u_1)u_1 + (b * u_2)u_2 = (u_1 u_1^T + u_2 u_2^T)b$
        i. $\hat{b} = UU^T b$ , which is linear transformation
            - matrix $U = [u_1 \;\; u_2]$
2. Application of orthogonal projection

a. SVN, encoder, ...

# Gram-Schmidt Orthogonalization & QR Factorization

▼ Keywords

  ▼ Gram-schmidt orthogonalization

  ▼ QR factorization

1. Gram-schmidt orthogonalization
    a. transform the two linearly independent vectors $V_1, V_2$ to orthonormal vectors $u_1, u_2$
        i. the span of the vectors $V$ and $u$ are the same
    b. how?
        i. normalize the vector $V_1$ to generate the vector $u_1$
        ii. compute $V_2 - u_1$ to make the vector $u_2$
        iii. normalize the vector $u_2$
2. QR factorization
    a. reproduce the vector $V_2$
    b. $A = QR$
        i. matrix $A$ is composed of the vectors $V_1, V_2$
        ii. matrix $R$ is a triangular matrix