Topic 2: Regression & Linear Least Squares

> - Linear Least Squares is a specific type of regression

Def: Regression

Regression refers to the collection of techniques for describing (capturing) dependencies between different data sets using "simple" functional relationship. We do this by specifying the functional relationship as:

$$y = f_\lambda(x) + r$$

- Note: sometimes, we denote r as ϵ

where:

- f is the functional relationship
- r is some kind of noise factor or secondary order effects not captured by the functional relationship
- λ: we use λ in f_λ to emphasize that when defining the functional form, we use **a family of functions** to describe the model. Then we convert the regression model to an optimization problem over the parameter space defined by λ. We then solve the optimization problem to find the specific functions (within the family) which "best" fits into the data we have at hand.
  - but what does "best" mean in here? $\rightarrow$ Here, "best" means that you have a measure of closness between function and data, usually expressd as a norm on **r**. In otherwords, we are minimizing the residual in a given norm.

In the context of linear algebra, we usually take our functional form to be a linear combination of vectors. This appears restrictive, however many non-linear problems can be reduced to linear problems by increasing the dimension of the independent variable space (example will be given shortly)

Linear Least Squares Regression

> - LLSP: Linear Least Squares Problem

Definition

Given $y \in \mathbb{R}^m, v_1, \dots v_n \in \mathbb{R}^m, m \geq n,$

- **Side Note:** $m = n$ means that the data can be fully captured by the relationship with the residual $r = 0$

find $\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n$ ($\alpha \neq \lambda$ showed before) such that $y \approx \alpha_1 v_1 + \dots + \alpha_n v_n$

Formally, LLSP is:

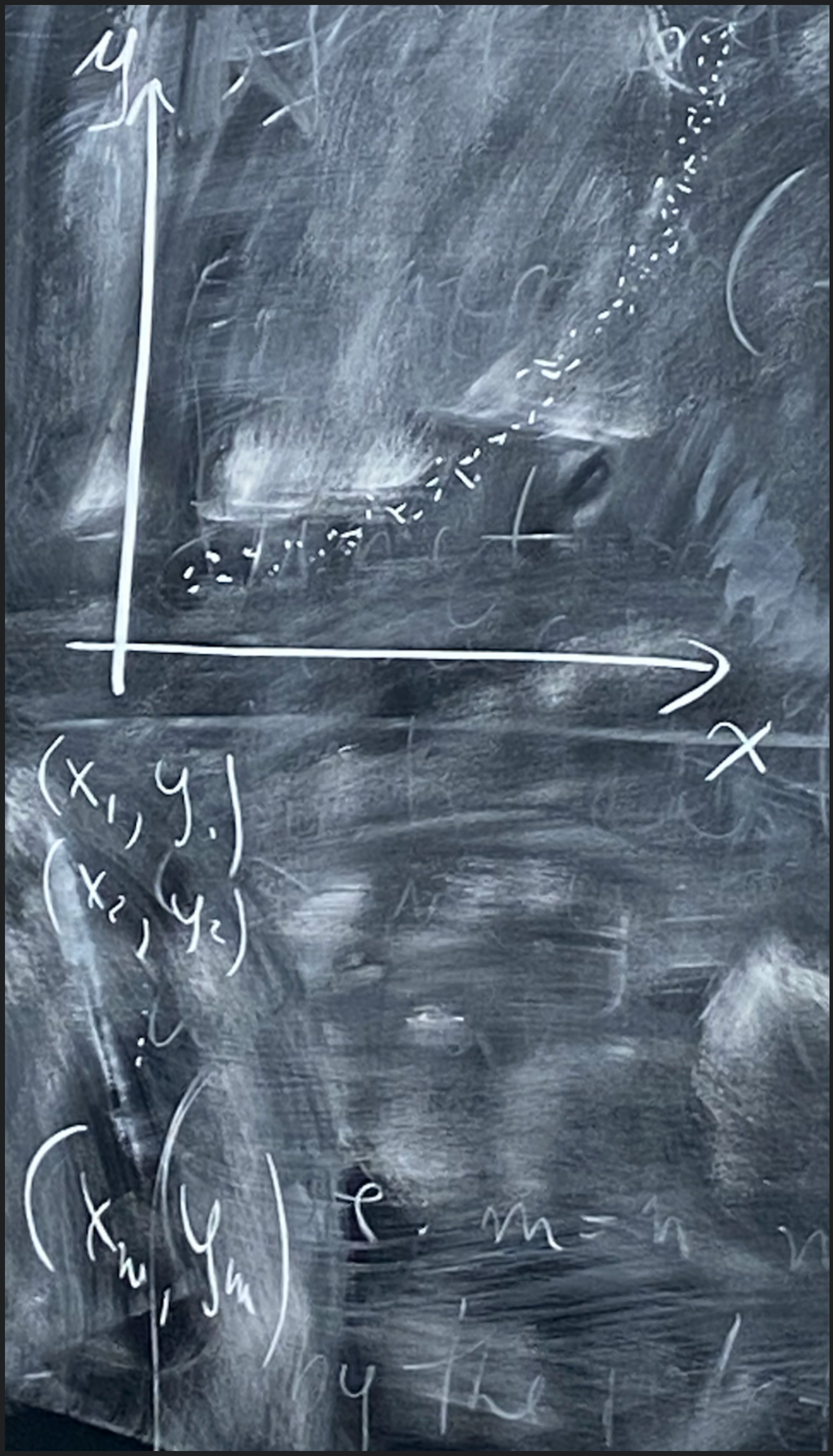
$$\min_{\alpha \in \mathbb{R}^n} ||y - A\alpha||_2$$

$$\text{where: } A = [v_1, \dots v_n], \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n$$

The overwhelming majority of all regressions are LLS regressions because:

1. Simple to formulate
2. Simple to solve
3. Many more sophisticated models do not perform much better

Schematic Example



Suppose in the picture above, you believe that this relationship is captured by a quadratic dependency. Then you model it as a parabola by stating:

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + r$$

In vector form, we can write:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \alpha_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \alpha_1 \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} + \alpha_2 \begin{pmatrix} x_1^2 \\ \vdots \\ x_m^2 \end{pmatrix} + \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix}$$

- where $y, x \in \mathbb{R}^m$

We can also write in matrix-vector form:

$$y = A\alpha + r$$

$$\text{where: } A = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{pmatrix}, \alpha = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

And: $v_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}, v_3 = \begin{pmatrix} x_1^2 \\ \vdots \\ x_m^2 \end{pmatrix}$

To solve, find $\min\limits_{\alpha \in \mathbb{R}^3} ||y - A\alpha||_2$

Existence and Uniquness of LLSP

Existence

The set $A\alpha, \alpha \in \mathbb{R}^n$ defines a subspace spanned by the column of A, and the problem $\min\limits_{\alpha \in \mathbb{R}^3} ||y - A\alpha||_2$ has a minimum defined by the projection of y onto that space. `(A2Q1)`

Uniqueness

1. If A is full-column rank, then the columns of A form a basis for $colspace(A)$ so the $proj_{colspace(A)}(y)$ is written uniquely in that basis.
2. If A is not full-column rank (i.e. have redundancy from basis point of view in the columns of A), then $proj_{colspace(A)}(y)$ can be written in infinitely many ways (i.e. infinitely many α 's achieve that projection)

Approches to Solve LLSP

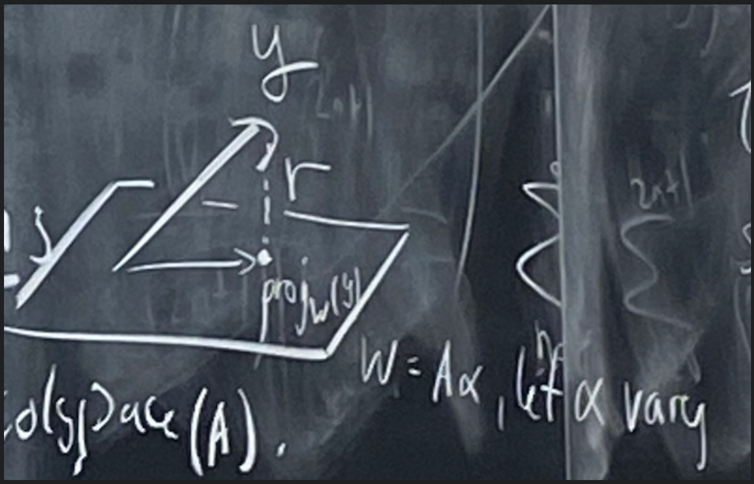
Recall LLSP:

$$\min_{\alpha \in \mathbb{R}^n} ||y - A\alpha||_2$$
$$\text{where: } A = [v_1, \dots v_n] \in \mathbb{R}^{m \times n}, m \geq n$$
$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n, y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m$$

This is the LLS Regression expressed as a minimization problem.

Let's look at techniques for solving this:

First Approach: Normal Equation



Use the fact that $r \perp W = colspace(A)$. This is expressed by saying that for every basis vector v of W, $r \perp v, v^T r = 0$

For the full column rank matrix A, the columns of A are a basis so we express this orthogonality condition as:

$$\left.\begin{array}{r} v_1^T r = 0 \\ v_2^T r = 0 \\ \vdots \\ v_n^T r = 0 \end{array}\right\} \implies A^T r = 0$$

This is the basic idea of normal equation.

Replace r by $y - A\alpha$ and rewrite as:

$$\begin{gathered} 0 = A^T(y - A\alpha) \\ 0 = A^T y - A^T A\alpha \\ A^T A = A^T y \end{gathered}$$

also known as the normal equations

Such rewriting reduces the original minimization problem to a linear system with $A^T A$ square

- shape of $A^T A = n \times m \cdot m \times n \to n \times n$

Note:

- this theory of closest vector using Inner Product expressions and orthogonality only applies to the 2-norm, i.e. theone induced by the I.P.

Here since A has rank n (which is the condition of full column rank & the fact that $m \geq n$), then $A^T A$ is invertible

- **Note:** such magical logic inferrence will be proved in `A4` using `SVD`

Thus the solution to the $LLSP$ is: $\alpha = (A^T A)^{-1}(A^T y)$

Second Approach: QR decomposition

This is a variant of the normal equations because we derive the solution using the same orthogonality condition, however instead of using the columns of A as a basis for the subspace $W = \{A\alpha : \alpha \in \mathbb{R}^n\}$, we use the columns of Q in the **reduced** QR decomposition, again assuming that A is full column rank.

we have:

$$
\begin{aligned}
0 &= Q^T r \\
&= Q^T (y - A\alpha) \\
&= Q^T (y - QR\alpha) \text{ Reduced QR decomposition} \\
&= Q^T y - Q^T QR\alpha \text{ Note: } (Q^T Q = I \text{, identity n x n }) \\
&= Q^T y - R\alpha
\end{aligned}
$$

Then:

$$R\alpha = Q^T y$$

This equation does not need matrix inversion to be solved because R is upper triangular so the solution is found by backward substituition.

Third Approach: Singular Value Decomposition

- TO BE SEEN LATER!

Comparing the 3 Approches

1. The normal equations work well when both following conditons hold:
    1. $m >> n$
    2. *condition number* of A is small (close to 1)
        - *condition number:* to be defined later. For now, just know it's the largest singular value over the smallest singular values

    when we have these 2 conditions, **approach 1** is good to use numerically & practically because it's a simple approach
2. When both m & n are large, **approch 2** is a good method numerically, more precise than **approch 1**

    Both **approch 1** & **approch 2** require A to be full column rank

3. **approch 3** is particularly good if A is rank-deficient (不足的) (i.e. $rank(A) < n$) or nearly so. There is one condition which we will make more precise later.

Concrete Example

Question:

Experimental data produces the following table:

t_i	y_i
1.00	1.84
1.10	1.96
1.30	2.21
1.50	2.45
1.90	2.94

You believe that this is a quadratic $y \approx \alpha_0 + \alpha_1 t + \alpha_2 t^2$, and you wish to find which quadratic fits this data optimally on the Least Squares sense.

Solution:

write as $LLSP$:

$$\begin{bmatrix} 1.84 \\ 1.96 \\ 2.21 \\ 2.45 \\ 2.94 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1.1 & 1.21 \\ 1 & 1.3 & 1.69 \\ 1 & 1.5 & 2.25 \\ 1 & 1.9 & 3.61 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$\Updownarrow$$

$$y = A\alpha$$

Let's solve by QR decomposition (in exam, use the fastest possible approach)

$$A = QR \text{ (reduced)}$$

$$Q = \begin{bmatrix} 0.4472 & -0.5031 & 0.4951 \\ 0.4472 & -0.3634 & 0.0724 \\ 0.4472 & -0.0839 & -0.4587 \\ 0.4472 & 0.1957 & -0.5768 \\ 0.4472 & 0.7547 & 0.4620 \end{bmatrix}$$

$$R = \begin{bmatrix} 2.2361 & 3.0411 & 4.3648 \\ 0 & 0.7155 & 2.0801 \\ 0 & 0 & 0.1909 \end{bmatrix}$$

Then solve $R\alpha = Q^T y$ by backwards substitution (use Matlab) to get:

$$\alpha = \begin{bmatrix} 0.6061 \\ 1.2387 \\ -0.0055 \end{bmatrix}$$

This completes the answer, the following statement is observation which is not part of the answer if not asked

Here (without going to statistical testing), you can observe that $\alpha_2 \approx 0$ so that your assumption of quadratic is like inacurate & a linear interpolation.