1. normal eq's
2. geometric interpretation of OLS.
3. hypothesis testing (under normality)

Normal equations:

$$\boxed{\hat{\beta}} \in \underset{\beta}{\operatorname{argmin}} \underbrace{\sum_{i=1}^n \frac{1}{2}(y_i - x_i^T \beta)^2}_{SSR(\beta)}$$

↑ OLS estimator

<u>Idea:</u> differentiate $SSR(\beta)$ & set derivative to zero

$$SSR(\beta) = \sum_{i=1}^n \underbrace{\frac{1}{2}(y_i - x_i^T \beta)^2}_{SSR_i(\beta)}$$

$$SSR_i(\beta) = \frac{1}{2}(\underbrace{y_i}_{1\times 1} - \underbrace{x_i^T}_{1\times p} \underbrace{\beta}_{p\times 1})^2$$

$$\nabla SSR(\beta) = \sum_{i=1}^n \nabla SSR_i(\beta)$$

$$\nabla SSR_i(\beta) = \nabla_\beta \left\{ \frac{1}{2}(y_i - x_i^T \beta)^2 \right\}$$

$$= \underbrace{x_i}_{p\times 1} \underbrace{(x_i^T \beta - y_i)}_{1\times 1}$$

$$\nabla SSR_i(\beta) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} SSR_i(\beta) \\ \vdots \\ \frac{\partial}{\partial \beta_p} SSR(\beta) \end{bmatrix}$$

$$\frac{\partial}{\partial \beta_1} SSR_i(\beta) = \frac{\partial}{\partial \beta_1} \left\{ \frac{1}{2}(y_i - \underbrace{x_i^T \beta}_{=\sum_{j=1}^p x_{ij}\beta_j})^2 \right\}$$

$$= \frac{\partial}{\partial \beta_1} \left\{ \frac{1}{2}\left(y_i - \sum_{j=1}^p x_{ij} \beta_j\right)^2 \right\}$$

$$= -(y_i - x_i^T \beta) x_{i,1}$$

$$\nabla SSR_i(\beta) = \begin{bmatrix} x_{i,1}(x_i^T \beta - y_i) \\ \vdots \\ x_{i,p}(x_i^T \beta - y_i) \end{bmatrix} = x_i (x_i^T \beta - y_i)$$

$$\nabla SSR(\beta) = \sum_{i=1}^n SSR_i(\beta) = \sum_{i=1}^n x_i (x_i^T \beta - y_i)$$

Define
$$X = \begin{bmatrix} - & x_1^T & - \\ & \vdots & \\ - & x_n^T & - \end{bmatrix} \Big\} n \text{ rows} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \Big\} n \text{ rows}$$
p-cols

---

$$\nabla SSR(\beta) = \sum_{i=1}^n x_i (x_i^T \beta - y_i)$$
$$= \underbrace{\begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}}_{n \text{ cols}} \begin{bmatrix} x_1^T\beta - y_1 \\ \vdots \\ x_n^T\beta - y_n \end{bmatrix}$$
p rows

$$\begin{bmatrix} x_1^T\beta \\ \vdots \\ x_n^T\beta \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$
$$X\beta \qquad\qquad y$$
$$\begin{bmatrix} - & x_1^T & - \\ & \vdots & \\ - & x_n^T & - \end{bmatrix} \beta$$

$$= \underbrace{X^T}_{p\times n} (\underset{n\times p}{X} \underset{p\times 1}{\beta} - \underset{n\times 1}{y})$$

---

Aside: slick "trace trick" in notes to compute $\nabla SSR(\beta)$

---

Summary: $\nabla SSR(\beta) = X^T(X\beta - y)$

Normal equations: $X^T(X\hat\beta - y) = 0$

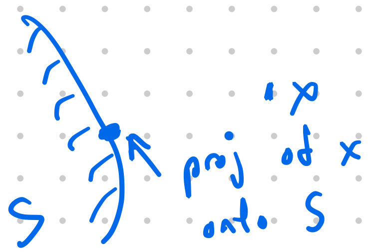
$$X^T X \hat\beta = X^T y$$

This is a system of linear eq's in $\beta$

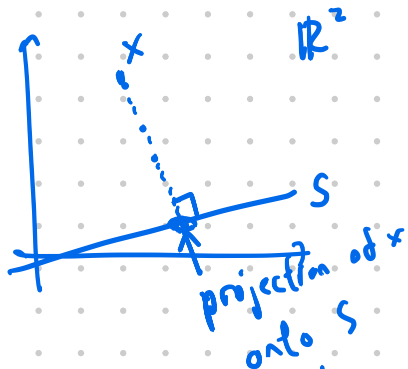
Assume $n \geq p$, and $X$ has full col rank, then $X^T X\hat\beta = X^T y$ always has a unique solution.

## Geometric interpretation of OLS

Def (projection): A pt $y$ is the projection of another pt $x$ onto a set $S \subset \mathbb{R}^n$ iff
$$y \in \underset{z}{\arg\min} \; \|x - z\|_2 \quad \text{s.t. } z \in S.$$




Fact (projection): If $S$ is subspace, then $y$ is the projection of $x$ onto $S$ iff $x - y \perp S$.




Recall from linear alg that the projection onto a subspace is given by a matrix $P$ s.t.
$$P^2 = P \;\; \& \;\; P = P^T.$$

$\|x\|_2 = (x^T x)^{\frac{1}{2}}$

There is a special $P \in \mathbb{R}^{n \times n}$ s.t. $Px \in \underset{z}{\arg\min} \; \|x - z\|_2$ s.t. $z \in S$

Check that $Px$ satisfies projection thm:

$$(x - Px) \perp z \text{ for any } z \in S$$

Here $P$ is the projector onto $S$

$$(x - Px)^T z = x^T z - (Px)^T z$$

$= x^T z - x^T P z \qquad (P = P^T)$

$= x^T z - x^T z \qquad (Pz = z$ because $z \in S)$

---

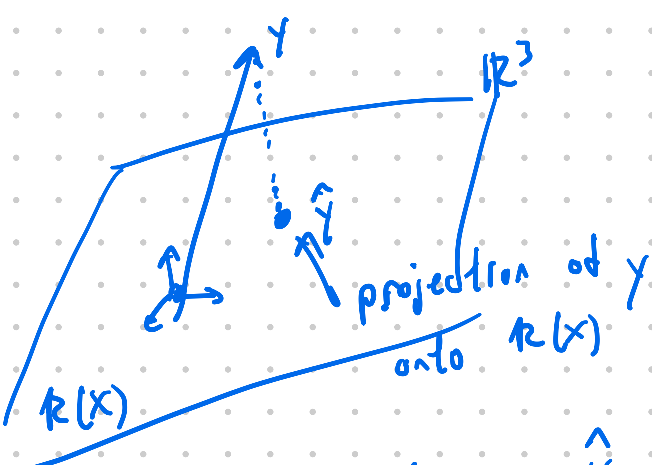
OLS: $\hat{\beta} \in \arg\min_{\beta} \underbrace{\sum_{i=1}^{n} \frac{1}{2}(y_i - x_i^T \beta)^2}_{=\ \frac{1}{2}(y - X\beta)^T (y - X\beta)}$

$$\frac{1}{2}(y - X\beta)^T (y - X\beta) = \frac{1}{2}\begin{bmatrix} y_1 - x_1^T\beta \\ \vdots \\ y_n - x_n^T\beta \end{bmatrix}^T \begin{bmatrix} y_1 - x_1^T\beta \\ \vdots \\ y_n - x_n^T\beta \end{bmatrix}$$

$$\hat{\beta} \in \arg\min_{\beta} \frac{1}{2}\|y - \underbrace{\boxed{X\beta}}_{z}\|_2^2$$

$$\hat{\beta} \in \left\{ \begin{array}{ll} \arg\min_{\beta} & \frac{1}{2}\|y - z\|_2^2 \\ \text{s.t.} & z = X\beta \end{array} \right\} = \left\{ \begin{array}{ll} \arg\min_{z} & \frac{1}{2}\|y - z\|_2^2 \\ \text{s.t.} & z \in R(X) \end{array} \right\}$$

$$= \left\{ \begin{array}{ll} \arg\min_{z} & \|y - z\|_2 \\ \text{s.t.} & z \in R(X) \end{array} \right\}$$




By projection fact, $y - \underbrace{\hat{y}}_{= X\hat{\beta}} \perp R(X)$

$$y - X\hat{\beta} \perp R(X)$$

Claim: $y - \hat{y} \perp R(X)$ iff $y - \hat{y} \perp$ cols of $X$

$y - \hat{y} \perp$ cols of $X$

$$(y-\hat{y})^T X = 0$$

$$(y-\hat{y})^T \underbrace{\begin{bmatrix} | & & | \\ x_{:,1} & \cdots & x_{:,p} \\ | & & | \end{bmatrix}}_{X} = 0$$

$$X^T(y-\hat{y}) = 0$$

$$X^T(y - X\hat{\beta}) = 0$$

---

solving the normal equations

1. compute QR factorization of $X$
2. form the product $v = Q^T y$
3. solve $R\hat{\beta} = v$ for $\hat{\beta}$

normal eq's: $X^T X \hat{\beta} = X^T y$

$$(QR)^T (QR)\hat{\beta} = (QR)^T y \qquad \left(\begin{array}{c}\text{plug in} \\ X = QR\end{array}\right)$$

$$R^T \underbrace{Q^T Q}_{I_p} R\hat{\beta} = R^T Q^T y$$

$$R^T R \hat{\beta} = R^T Q^T y \qquad (Q^T Q = I_p)$$

$$R^{-T} R^T R \hat{\beta} = R^{-T} R^T Q^T y \qquad (R^{-T} R^T = I_p)$$

$$R\hat{\beta} = Q^T y$$