Math 342W/642/742W

Recitation – Day #10 (3.13.25)

I. The Projection Matrix H with QR-Decomposition

(i) What is the matrix Q that comes from decomposing our model/design matrix X into its QR Factorization?

Given an $n \times (p+1)$ design matrix X of (full) rank equal to $p+1$, the matrix Q is an $n \times (p+1)$ orthogonal matrix with orthonormal column vectors that are found after the columns of X,which are a basis for the column space of X, go through the Gram-Schmidt orthogonalization process.

(ii) What do we know about the colsp[Q] and the colsp[X]?

$$\text{colsp}[Q] = \text{colsp}[X]$$

(iii) What is an equivalent way to define our projection matrix H onto the column space of X?

$$H = QQ^T$$

(iv) How do we express our vector of predicted responses $\hat{\boldsymbol{y}}$ with H? now with Q? with projections?

$$\hat{\boldsymbol{y}} = H\boldsymbol{y} = QQ^T\boldsymbol{y} = \sum_{j=0}^{p} \text{proj}_{q_j}(\boldsymbol{y})$$

Note: The last equality is due to the orthogonality of the $\boldsymbol{q}_j$'s.

(v) How does the pythagorean theorem help us with understanding more about $\|\hat{\boldsymbol{y}}\|^2$?

Applying the pythagorean theorem to the expression in part (iv) we get:

$$\begin{aligned}
\|\hat{\boldsymbol{y}}\|^2 &= \sum_{j=0}^{p} \left\|\text{proj}_{\boldsymbol{q}_j}(\boldsymbol{y})\right\|^2 \\
&= \left\|\text{proj}_{\boldsymbol{q}_0}(\boldsymbol{y})\right\|^2 + \sum_{j=1}^{p} \left\|\text{proj}_{\boldsymbol{q}_j}(\boldsymbol{y})\right\|^2 \\
\implies \|\hat{\boldsymbol{y}}\|^2 - n\bar{y}^2 &= \sum_{j=1}^{p} \left\|\text{proj}_{\boldsymbol{q}_j}(\boldsymbol{y})\right\|^2
\end{aligned}$$

II. Insights of OLS through QR-Decomposition

(i) What makes $\left\|\text{proj}_{\boldsymbol{q}_0}(\boldsymbol{y})\right\|^2$ special?

$$\left\|\text{proj}_{\boldsymbol{q}_0}(\boldsymbol{y})\right\|^2 = \left\|\text{proj}_{\mathbf{1}_n}(\boldsymbol{y})\right\|^2 = \left\|\frac{\mathbf{1}_n \cdot \mathbf{1}_n^T}{\|\mathbf{1}_n\|^2}\boldsymbol{y}\right\|^2 = \left\|\frac{1}{n}\begin{bmatrix}1 & \cdots & 1\\ \vdots & \ddots & \vdots\\ 1 & \cdots & 1\end{bmatrix}\boldsymbol{y}\right\|^2 = \|\bar{y}\cdot\mathbf{1}_n\|^2$$
$$= (\bar{y}\mathbf{1}_n)^T(\bar{y}\cdot\mathbf{1}_n) = \bar{y}^2\cdot\mathbf{1}_n^T\mathbf{1}_n = n\bar{y}^2$$

(ii) How is SSR related to the projection of $\boldsymbol{y}$ along the orthogonal columns of Q?

$$\begin{aligned}\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 &= (\bar{y} - \bar{y}\cdot\mathbf{1}_n)^T(\bar{y} - \bar{y}\cdot\mathbf{1}_n)\\ &= \hat{\boldsymbol{y}}^T\hat{\boldsymbol{y}} - \bar{y}\cdot\mathbf{1}_n^T\hat{\boldsymbol{y}} - \bar{y}\cdot\hat{\boldsymbol{y}}^T\cdot\mathbf{1}_n + \bar{y}^2\cdot\mathbf{1}_n^T\mathbf{1}_n\\ &= \|\hat{\boldsymbol{y}}\|^2 - 2\bar{y}\cdot\hat{\boldsymbol{y}}^T\mathbf{1}_n + n\bar{y}^2\\ &= \|\hat{\boldsymbol{y}}\|^2 - 2n\bar{y}^2 + n\bar{y}^2\\ &= \|\hat{\boldsymbol{y}}\|^2 - n\bar{y}^2\\ &= \sum_{j=1}^{p}\left\|\text{proj}_{q_j}(\boldsymbol{y})\right\|^2 && \textit{from Section I. part (v)}\end{aligned}$$

Note: $\hat{\boldsymbol{y}}^T\mathbf{1}_n = (H\boldsymbol{y})^T\mathbf{1}_n = \boldsymbol{y}^T H^T\mathbf{1}_n = \boldsymbol{y}^T H\mathbf{1}_n = \boldsymbol{y}^T\mathbf{1}_n = n\bar{y}$

(iii) What insight does this all give us with including a new feature(s) $\boldsymbol{x}_{\cdot\star}$ to the model/design matrix X under OLS?

As long as you add a new feature $\boldsymbol{x}_{\cdot\star}$ such that $X_\star = [X \vdots \boldsymbol{x}_{\cdot\star}]$ is still full rank, then $Q_\star = [Q \vdots \boldsymbol{q}_{\cdot\star}]$ and $\text{SSR}_\star = \text{SSR} + \underbrace{\|\text{proj}_{\boldsymbol{q}_\star}\|^2}_{>0} \implies \text{SSR}_\star > \text{SSR}$.

If we were to add $n-(p+1)$ columns to X due to new independent features, we would have a new matrix $X_\star$ of full rank and would have the following:

$$\hat{\boldsymbol{y}}_\star = H_\star\boldsymbol{y} = X_\star\left(X_\star^T X_\star\right)^{-1}X_\star^T = \underbrace{X_\star X_\star^{-1}}_{I}\underbrace{(X_\star^{-1})X_\star^T}_{I}\boldsymbol{y} = \boldsymbol{y}$$
$$\implies \hat{\boldsymbol{y}} = \boldsymbol{y} \implies \text{SSE} = 0 \implies \text{RMSE} = 0 \implies R^2 = 1$$
$$\implies \text{performance metrics are “dishonest” due to overfitting occuring}$$

(iv) What is the relationship between R^2, SSR, SSE and RMSE?

Recall that, SST = SSR + SSE. Let SST be a fixed quantity.

$$\begin{array}{ccc}\text{SSR}\uparrow & = & \text{SSE}\downarrow\\ \| & & \|\\ R^2\uparrow & = & \text{RMSE}\downarrow\end{array}$$