IML

הגשה 2

רוני פרלס

312263032

(1) שאלה 1 (1) צ"ל: $ker(X) = ker(X^TX)$

נראה הכלה דו כיוונית: כל u שנמצא בגרעין של X נמצא בגרעין של X^TX

$\underline{u \in ker(X)}$: , $u \neq 0$ צ"ל:

$$X^TXu = X^T(Xu) \underset{\substack{\downarrow \\ u \in ker(X)}}{=} X^T \cdot \vec{0} = \vec{0} \Rightarrow u \in ker(X^TX)$$

$\underline{u \in ker(X^TX)}$: , $u \neq 0$ צ"ל:

$$X^TXu = 0 \Rightarrow (u^TX^T)Xu = 0 \underset{\substack{\downarrow \\ \text{מכפלה פנימית}}}{\Rightarrow} (Xu)^T(Xu)$$

$$\Rightarrow Xu = 0 \Rightarrow u \in ker(X)$$

הראנו הכלה דו כיוונית לכן $ker(X) = ker(X^TX)$ מ.ש.ל

(2) A מטריצה כלשהי צ"ל: $Im(A^T) = ker(A)^{\perp}$

נראה הכלה דו כיוונית:

יהי $u_1 \in Im(A^T) \Leftarrow$ $\exists x_1 \in \mathbb{R}^n$ s.t $A^Tx_1 = u_1$

יהי $u_2 \in ker(A) \Leftarrow$ $Au_2 = 0$

נראה $u_1 \in ker(A)^{\perp}$ כלומר $\langle u_1, u_2 \rangle = 0$ לכל $u_2 \in ker(A)$:

$$\langle u_1, u_2 \rangle = \langle A^Tx_1, u_2 \rangle = \langle x_1, Au_2 \rangle = \langle x_1, 0 \rangle = 0$$

לכן $u_1 \in ker(A)^{\perp} \Rightarrow Im(A^T) \subseteq ker(A)^{\perp}$

כיוון שני: יהי $x \in ker(A)^{\perp}$

נניח בשלילה ש $x \notin Im(A^T)$ אז קיים $Im(A^T)^{\perp} \ni x'$

כך ש $\langle x, x' \rangle \neq 0$ (*)

נסמן $Im(A) \ni u = A^TAx'$ אז

$$\langle Ax', Ax' \rangle = \langle x', A^TAx' \rangle = \langle x', u \rangle \underset{\substack{\downarrow \\ \text{אורתוגונליים}}}{=} 0 \Rightarrow Ax' = 0$$

$$\Downarrow$$

$$x' \in ker(A)$$

$$\Downarrow$$

$\langle x, x' \rangle = 0$ סתירה ל- (*), לכן $x \in Im(A^T) \Rightarrow ker(A)^{\perp} \subseteq Im(A^T)$

הראנו הכלה דו כיוונית לכן מתקיים שוויון מ.ש.ל

(ג) $y = Xw$ מערכת משוואות לינאריות לא הומוגניות

X ריבועית ולא הפיכה, צ"ל מערכת עם פתרונות

אם"מ $y \perp \ker(X^T)$

עבור מטריצה שאינה הפיכה קיימת ת"ל לכן יש לה אינסוף פתרונות או שלא קיימים לה פתרונות.

אם נניח כי קיימים אינסוף פתרונות, אם"מ יש ת"ל

$$\Updownarrow$$

(קיים פתרון) ∧ (יש ת"ל ב $y = Xw$)

$$\Updownarrow$$

קיים פתרון למערכת $y \in Im(X)$

$$\Updownarrow$$

משפט 2

$$y \in \ker(X^T)^{\perp}$$

$$\Updownarrow$$

$$y \perp \ker(X^T)$$

מ.ש.ל.

(ד) $X^T X w = X^T y$

נוכיח שיש פתרון יחיד אם $X^T X$ הפיכה, אחרת יש ∞ פתרונות.

תהי $X^T X$ לא הפיכה

משפט 1 $\ker(X^T X) = \ker(X)$

תהי $u \in \ker X^T \Leftrightarrow u \in \ker(X^T X) \Rightarrow X^T X u = 0$

$\langle X^T X u, y\rangle = \langle Xu, X^T y\rangle = 0 \Rightarrow y \perp \ker(X^T)$

לכן מ-3 עבור $X^T X$ לא הפיכה יש אינסוף פתרונות (אם"מ)

מכאן שיש פתרון יחיד אם היא הפיכה. מ.ש.ל.

(5)

$$P=\sum_{i=1}^{k} v_i v_i^T$$

$V \subseteq \mathbb{R}^d,\ \dim(V)=k$

(א) כל איבר הסכום הוא מכפלה וקטור באנספוז שלו
של מטריצה סימטרית ולכן הסכום גם הוא מטריצה סימטרית

(ב)

$$v_j \in v_1 \dots v_k,\quad P v_j = \sum_{i=1}^{k} v_i v_i^T v_j = \sum_{i=1}^{k} v_i \delta_{ji} = v_j$$

ע"ע הוא 1

(ג) $v \in V$ נוכל לכתוב: $v=\sum_{i=1}^{k} \alpha_i v_i,\ \alpha_i \in \mathbb{R}$ אז

$$Pv = P\sum_{i=1}^{k} \alpha_i v_i = \sum_{i=1}^{k} \alpha_i P v_i = \sum_{i=1}^{k} \alpha_i v_i = v$$

(ד) נניח $P=UDU^T$, D אלכסונית עם 1 ו-0 על האלכסון $\Leftarrow$ $D^2=D$

$$P^2=UDU^T\cdot UDU^T = UD^2U^T = UDU^T = P$$

(ה) צ"ל $(I-P)P=0$

$$(I-P)P = P^2-P = P-P=0$$

$S=\{(x_i,y_i)\}_{i=1}^{m}$ the ERM: $\hat{w}\in \underset{w\in\mathbb{R}^d}{\operatorname{argmin}} \|Xw-y\|^2$

$X=U\Sigma V^T$ SVD of X, $U_{m\times m}$ orthogonal, $\sigma_i=\Sigma_{ii}$

$X^{\dagger}=V\Sigma^{\dagger}U^T \qquad \Sigma^{\dagger}_{i,i}=\begin{cases}\sigma_i^{-1} & \sigma_i\neq 0\\ 0 & \sigma_i=0\end{cases}$

(6) X^TX הפיכה

$$XX^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T = UDU^T \qquad D=\Sigma\Sigma^T$$

$$I=UDU^TUD^{-1}U^T \Rightarrow (XX^T)^{-1}=UD^{-1}U^T$$

צ"ל נשים לב $(XX^T)^{-1}X=X^{T\dagger}$

$$(XX^T)^{-1}X = UD^{-1}U^TU\Sigma V^T = U(\Sigma\Sigma^T)^{-1}U^TU\Sigma V^T =$$
$$= U\Sigma^{T\dagger}V^T = (V\Sigma^{\dagger}U^T)^T = X^{\dagger T}$$

(7) צ"ל X^TX הפיכה אמ"ם $\mathrm{Span}\{x_1,\dots,x_m\}=\mathbb{R}^d$

$\Leftarrow$ $\mathrm{Span}\{x_1,\dots,x_m\}=\mathbb{R}^d$, ממשפט המימדים $\dim(\ker X)=0$ $\Leftarrow$ X^T הפיכה

כלומר יש d ערכים סינגולריים ולכן גם מטריצה סינגולרית

בגודל $d\times d$ לכן $XX^T=\Sigma\Sigma^T$

כלומר יש לה מטריצה אלכסונית הפיכה בגודל $d\times d$

כמו XX^T הפיכה.

$\Rightarrow$ XX^T הפיכה, $\Leftarrow$ $\mathrm{rank}(XX^T)=\mathrm{rank}(X)=d$

$$\Downarrow$$

$$\mathrm{Span}\{x_1,\dots,x_m\}=\mathbb{R}^d$$

מ.ש.ל

(8) X^TX לא הפיכה $\Leftarrow$ הרבה פתרונות

צ"ל $\hat{w}=X^+y$ הפתרון ~~...~~ שנבחר הוא בעל נורמה מינימלית

$\Leftarrow$ לכל פתרון אחר $\bar{w}$: $\|\hat{w}\|\le\|\bar{w}\|$

מהגדרת Σ^+ ניתן להבין שלכל $i\le r$

$\hat{w}=\bar{w}$. עבור $i>r$ מתקיים $\hat{w}=0$ ועבור פתרונות

שונים מתקיים:

$$\|\hat{w}\|^2=\sum_{i=0}^{r}\hat{w}_i^2+\sum_{i=r+1}^{d}0=\sum_{i=0}^{r}\hat{w}_i^2=\sum_{i=0}^{r}\bar{w}_i^2=$$

$$=\sum_{i=0}^{r}\bar{w}_i^2+\sum_{i=r+1}^{d}0\le\sum_{i=0}^{r}\bar{w}_i^2+\sum_{i=r+1}^{d}\bar{w}_i^2=\|\bar{w}\|^2$$

$$\Uparrow$$

$$\bar{w}_i^2\ge 0\quad\forall i$$

מ.ש.ל

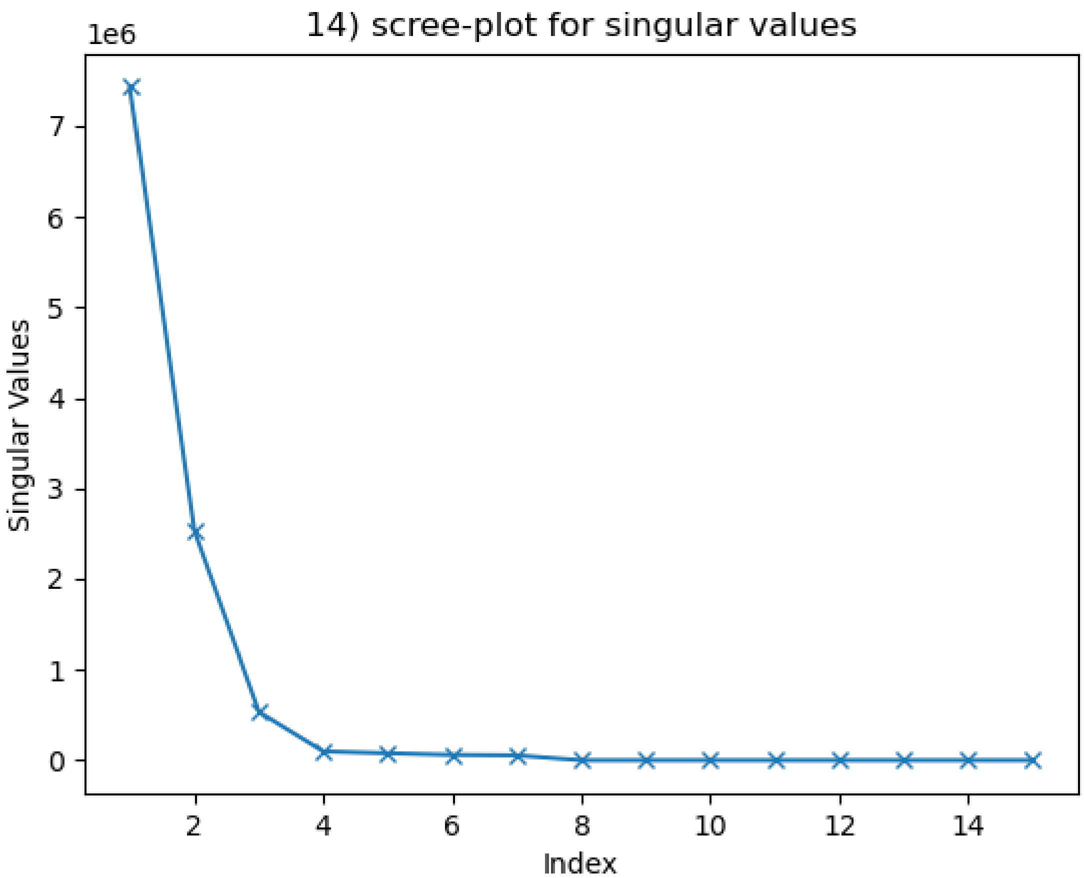
Answers to practical questions

Q13:

I found the following features to be categorical

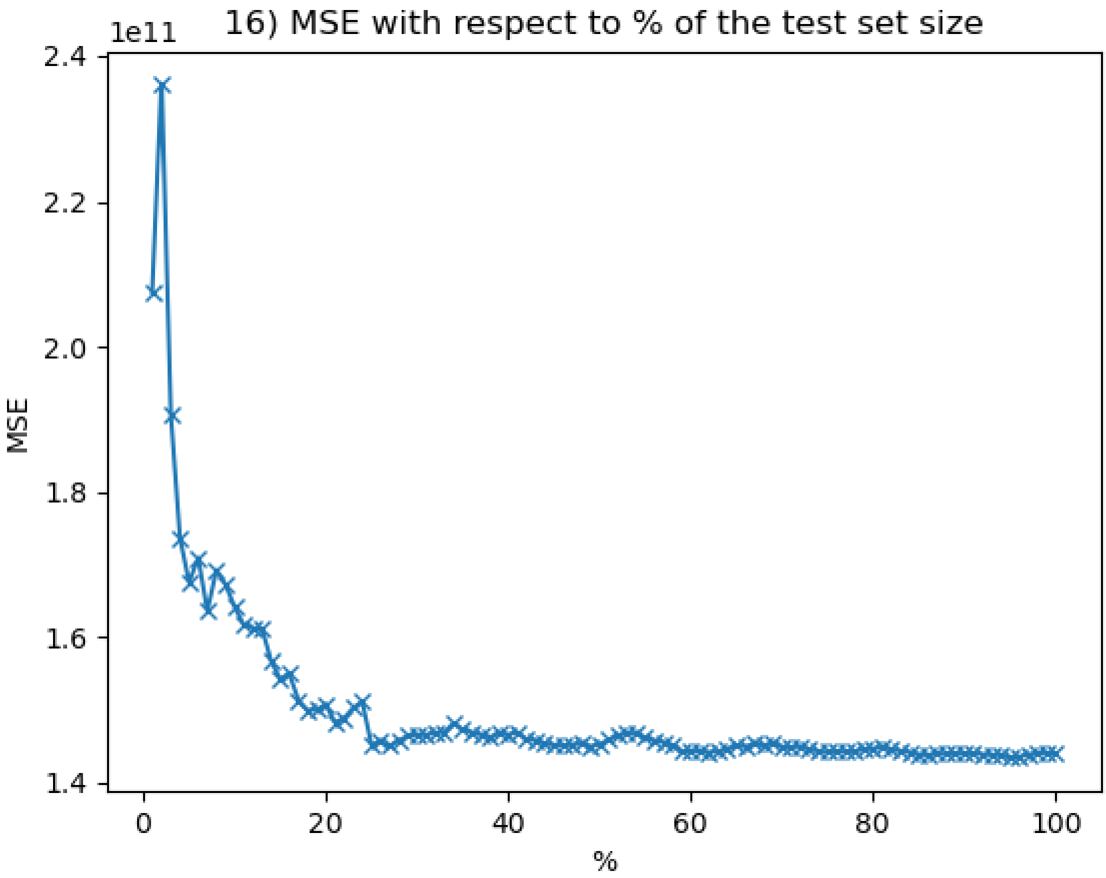
- **ID**: a unique number
- **Longitude/Latitude**: can be ordered but the order does not have meaning
- **Date**: since it's has meaning only over large period of time (years), and the only years available are 2014/2015, we can treat it as categorical data.

Q15:



The matrix is singular since it has values that are equal to 0
This means that the matrix is not invertible

Q16:

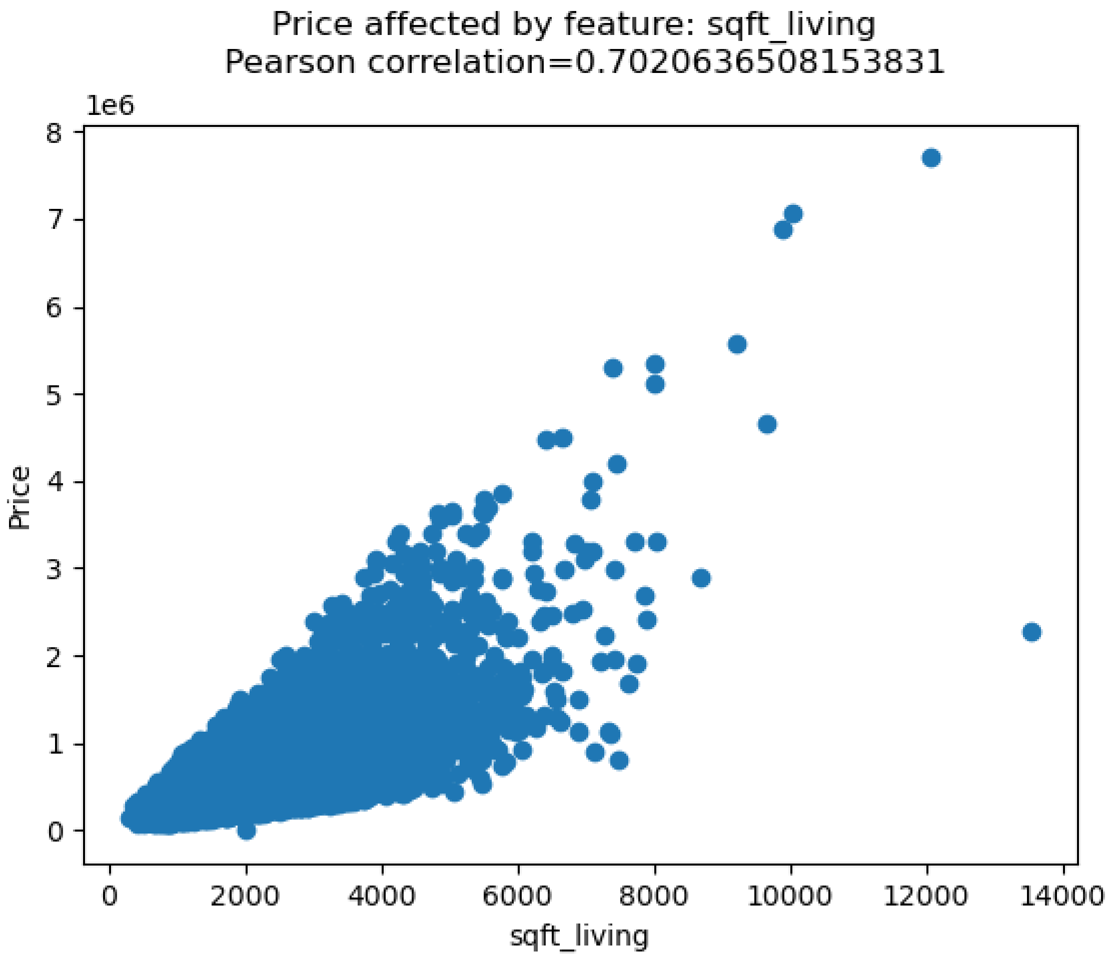


We can see that the MSE decreases with respect to decreasing the percentage of the test set size (increasing the train set size)

We can also see that increasing the percentage over ~25% does not change the does not change the MSE very much and it stays around a constant value.

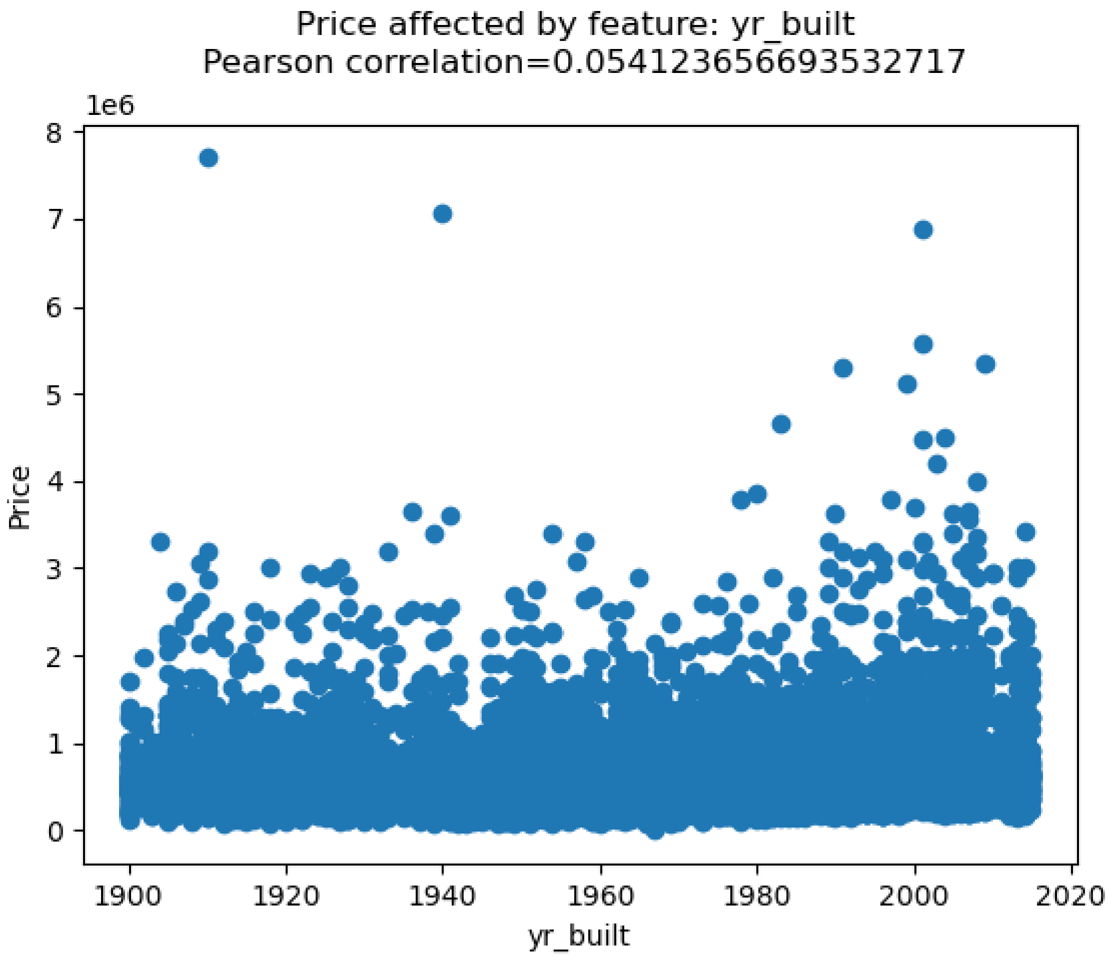
Q17:

Beneficial for the model:



We can see high correlation: the price increases when the as the size of the living room increases. Also the Pearson correlation is relatively high.
From this we can conclude that this type of data is beneficial for the model.

Not beneficial for the model:



We can see low correlation: the price does not have a particular trend as the size of the living room changes. Also the Pearson correlation is relatively low.
From this we can conclude that this type of data is not very beneficial for the model.