למידה חישובית בייסיאנית, סמסטר א׳ 2022-2023

תרגיל 2: רגרסיה לינארית בייסיאנית

רינה קרנאוך 319132353

הסטיקר שילווה אותנו בתרגיל זה:

(זה חתול ישן ובריא ונושם)

1. <u>חלק תאורטי</u>

1.1. <u>משערכים</u>

בהינתן θ, α, γ הפרמטרים ו $\eta \sim \mathcal{N}(0, I\sigma^2)$ עבור σ^2 שאיננה ידועה. נתבונן על 3 דרכים של פרמטריזציה של פולינומים מדרגה $k \in \mathbb{N}$.

i. $y_\theta(x) = \sum_{n=1}^{k} \theta_n x^n + \eta$

ii. $y_\alpha(x) = \sum_{n=1}^{k} (10\alpha_n) x^n + \eta$

iii. $y_\gamma(x) = \sum_{n=1}^{k} (\gamma_n)^3 x^n + \eta$

א. נניח ואנו יודעים לחשב את $p_\theta(\theta|\mathcal{D})$, מה יהיו ההתפלגויות הבאות (ביחס להתפלגות $p_\theta(\theta|\mathcal{D})$):

לפני הסעיפים, נרצה לנתח את ה$posterior$ הנתון לנו:

$$y_\theta(x) = \sum_{n=1}^{k} \theta_n x^n + \eta$$

נוכל לסמן את הסכימה בתור מכפלת מטריצות:

$$\theta = \begin{pmatrix} \theta_1 \\ \dots \\ \theta_k \end{pmatrix}, H = [h_1(x), h_2(x), \dots, h_k(x)]\ s.t\ \forall i \in [k] : h_i(x) = x^i$$

ונקבל:

$$y_\theta(x) = H\theta + \eta$$

a. $p_\alpha(\alpha|\mathcal{D})$ – נכתוב את $y_\alpha(x)$ בצורה שונה (בעזרת מטריצה H):

$$y_\alpha(x) = 10H\alpha + \eta$$

אנו יודעים מהתרגול הראשון, שעבור פונקציה $f_1 : \theta \to \alpha$ כך ש $f_1(\theta) = \alpha$, כך נקבל:

$$p_\alpha(\alpha|\mathcal{D}) = \left| \frac{\partial f_1^{-1}(\alpha)}{\partial \alpha} \right| p_\theta(f_1^{-1}(\alpha)|\mathcal{D})$$

ומשם נוכל לקבל את:

$$\mathbb{P}_\alpha(\alpha|\mathcal{D}) = \int_{-\infty}^{\alpha} p_\alpha(\tilde{\alpha}|\mathcal{D}) d\tilde{\alpha}$$

כעת נרצה למצוא את הפונקציה f_1 כנ״ל, וזאת בעזרת y:

$$y_{f_1(\theta)}(x) = y_\alpha(x) \rightarrow 10H\underbrace{f_1(\theta)}_{\alpha} + \eta = H\alpha + \eta$$

לכן:

$$f_1(\theta) = \underbrace{\frac{1}{10}\theta = \alpha}_{\substack{\theta=10\alpha \\ \alpha=\frac{1}{10}\theta}} \rightarrow f_1^{-1}(\alpha) = 10\alpha$$

ונקבל:

$$p_\alpha(\alpha|\mathcal{D}) = \left|\frac{\partial f_1^{-1}(\alpha)}{\partial \alpha}\right| p_\theta(f_1^{-1}(\alpha)|\mathcal{D})$$

$$= \left\| \begin{bmatrix} \frac{\partial\left(f_1^{-1}(\alpha)\right)_1}{\partial \alpha_1} & \cdots & \frac{\partial\left(f_1^{-1}(\alpha)\right)_k}{\partial \alpha_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial\left(f_1^{-1}(\alpha)\right)_1}{\partial \alpha_k} & \cdots & \frac{\partial\left(f_1^{-1}(\gamma)\right)_k}{\partial \alpha_k} \end{bmatrix} \right\| p_\theta(10\alpha|\mathcal{D})$$

$$= \left|10\begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}\right| = 10^k p_\theta(10\alpha|\mathcal{D})$$

b. $p_\gamma(\gamma|\mathcal{D})$ – נכתוב את $y_\gamma(x)$ בצורה שונה (בעזרת מטריצה H):

נגדיר:

$$\gamma = [\gamma_1, \gamma_2, \dots, \gamma_k]^T, \gamma^3 = [\gamma_1^3, \gamma_2^3, \dots, \gamma_k^3]^T$$

$$y_\gamma(\gamma) = H\gamma^3 + \eta$$

כעת נרצה למצוא פונקציה $f_2: \theta \rightarrow \gamma$ כך שנקבל $f_2(\theta) = \gamma$ כאשר נצטרך ש:

$$y_{f_2(\theta)}(x) = y_\gamma(x)$$

כלומר:

$$H\underbrace{\left(f_2(\theta)\right)^3}_{\gamma^3} + \eta = H\theta + \eta \rightarrow f_2(\theta) = \theta^{-3}$$

כלומר קיבלנו הגדרה עבור הפונקציה f_2^{-1}:

$$f_2^{-1}(x) = x^3$$

כאשר נגדיר שעבור ווקטור v:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_k \end{bmatrix} \rightarrow v^p = \begin{bmatrix} v_1^p \\ v_2^p \\ \dots \\ v_k^p \end{bmatrix}$$

נקבל

$$p_\gamma(\gamma|\mathcal{D}) = \left|\frac{\partial f_2^{-1}(\gamma)}{\partial \gamma}\right| \cdot p_\theta(f_2^{-1}(\gamma)|\mathcal{D})$$

$$= \left\|\begin{bmatrix} \frac{\partial\left(f_2^{-1}(\gamma)\right)_1}{\partial \gamma_1} & \cdots & \frac{\partial\left(f_2^{-1}(\gamma)\right)_k}{\partial \gamma_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial\left(f_2^{-1}(\gamma)\right)_1}{\partial \gamma_k} & \cdots & \frac{\partial\left(f_2^{-1}(\gamma)\right)_k}{\partial \gamma_k} \end{bmatrix}\right\| \cdot p_\theta(\gamma^3|\mathcal{D})$$

נקבל:

$$\begin{bmatrix} \frac{\partial\left(f_2^{-1}(\gamma)\right)_1}{\partial \gamma_1} & \cdots & \frac{\partial\left(f_2^{-1}(\gamma)\right)_k}{\partial \gamma_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial\left(f_2^{-1}(\gamma)\right)_1}{\partial \gamma_k} & \cdots & \frac{\partial\left(f_2^{-1}(\gamma)\right)_k}{\partial \gamma_k} \end{bmatrix} = \begin{bmatrix} 3\gamma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 3\gamma_k^2 \end{bmatrix}$$

ולכן:

$$\det\left(\begin{bmatrix} 3\gamma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 3\gamma_k^2 \end{bmatrix}\right) = \prod_{i=1}^{k} 3\gamma_i^2 = 3^k \prod_{i=1}^{k} \gamma_i^2$$

ולכן:

$$p_\gamma(\gamma|\mathcal{D}) = \left(3^k \prod_{i=1}^{k} \gamma_i^2\right) \cdot p_\theta(\gamma^3|\mathcal{D})$$

ב. נגדיר $k = 1$ כלומר נקבל ש$\theta, \alpha, \gamma \in \mathbb{R}$ מה גם:

$$y_\theta(x) = \theta x + \eta, y_\alpha(x) = 10\alpha x + \eta, y_\gamma(x) = \gamma^3 x + \eta$$

נגדיר $\hat{\theta}$ להיות $MMSE\ estimator$ של θ. נזכיר שמדובר ב $\mathbb{E}[\theta|\mathcal{D}]$ נשאלת השאלה האם הפתרונות $y_{\hat{\alpha}}(x), y_{\hat{\gamma}}(x)$ בהכרח יהיו בעלי אותו ערך עבור כל $\mathcal{D}, x$. כעת:

$$y_{\hat{\theta}}(x) = \hat{\theta} x + \eta, \qquad y_{\hat{\alpha}}(x) = 10\hat{\alpha} x + \eta, \qquad y_{\hat{\gamma}}(x) = \hat{\gamma}^3 x + \eta$$

כעת:

$$\hat{\alpha} = \mathbb{E}[\alpha|\mathcal{D}] = \int_{-\infty}^{\infty} p_\alpha(\tilde{\alpha}|\mathcal{D}) \cdot \tilde{\alpha} d\tilde{\alpha} = \int_{-\infty}^{\infty} p_\theta(10\tilde{\alpha}|\mathcal{D}) \cdot 10 \cdot \tilde{\alpha} d\tilde{\theta}$$

$$\hat{\gamma} = \mathbb{E}[\gamma|\mathcal{D}] = \int_{-\infty}^{\infty} p_\gamma(\tilde{\gamma}|\mathcal{D}) \cdot \tilde{\gamma} d\tilde{\gamma} = \int_{-\infty}^{\infty} p_\theta(\tilde{\gamma}^3|\mathcal{D}) \cdot 3\tilde{\gamma}^2 \cdot \tilde{\theta} d\tilde{\theta}$$

אנו יודעים ש $\hat{\theta}$ ימזער את $\mathbb{E}[\theta|\mathcal{D}] = \int_{-\infty}^{\infty} p_\theta(\tilde{\theta}|\mathcal{D}) \tilde{\theta} d\tilde{d}$, עבור כל ערך $\tilde{\theta}$.

אנו יודעים שעבור כל θ, γ:

$$\gamma^3 = \theta, 10\alpha = \theta$$

כעת נרצה לבדוק אם מתקיים שהערכים הממוזערים גם כן מקיימים:

$$\hat{\gamma}^3 = \hat{\theta}, 10\hat{\alpha} = \hat{\theta}$$

והרי:

$$\hat{\alpha} = \int_{-\infty}^{\infty} p_\theta\left(\underbrace{10\tilde{\alpha}}_{\tilde{\theta}}\middle|\mathcal{D}\right)\cdot 10\cdot\tilde{\alpha}d\tilde{\theta} = 10\underbrace{\int_{-\infty}^{\infty} p_\theta(\tilde{\theta}|\mathcal{D})\tilde{\theta}d\tilde{\theta}}_{\hat{\theta}} = 10\hat{\theta}$$

כלומר קיים שוויון בין הממזערים כנ"ל.

$$\hat{\gamma} = \int_{-\infty}^{\infty} p_\theta(\tilde{\gamma}^3|\mathcal{D})\cdot 3\tilde{\gamma}^2\cdot\tilde{\theta}d\tilde{\theta} = 3\int_{-\infty}^{\infty} p_\theta\left(\underbrace{\tilde{\gamma}^3}_{\tilde{\theta}}|\mathcal{D}\right)\cdot\underbrace{\tilde{\gamma}^2}_{\tilde{\theta}}\cdot\tilde{\theta}d\tilde{\theta} = 3\int_{-\infty}^{\infty} p_\theta(\theta|\mathcal{D})\cdot\tilde{\theta}^3d\tilde{\theta}$$

אבל ערך זה **איננו שווה ל** $\hat{\boldsymbol{\theta}} = \int_{-\infty}^{\infty} p_\theta(\theta|\mathcal{D})\cdot\tilde{\theta}d\tilde{\theta}$, על כן:

$$\hat{\gamma} \neq \hat{\theta}$$

1.2. <u>רגרסיה לינארית בייסיאנית</u>

בהינתן 2 מדגמים $\mathcal{D}_1, \mathcal{D}_2$ ובעיית רגרסיה. בשני המדגמים זוגות (x_i, y_i) ואנו מניחים שעבור שני המדגמים:

$$y_i = \sum_{n=1}^{d}\theta_n h_n(x_i) + \eta$$

עבור קבוצת פונקציות בסיס $\{h_i(\cdot)\}_{i\in[d]}$, כאשר $\eta\sim\mathcal{N}(0, I\sigma^2)$. נניח $prior$ גאוסיאני עבור θ כלומר:

$$\theta\sim\mathcal{N}(\mu,\Sigma)$$

i. נרצה להוכיח כי מתקיים:

$$p(\theta, D_1, D_2) = p(\theta)p(D_1|\theta)p(D_2|\theta)$$

והרי מהגדרת התפלגות משותפת נוכל לרשום:

$$p(\theta, D_1, D_2) = p(\theta)\cdot\underbrace{p(D_1, D_2|\theta)}_{\forall d_1,d_2\in D_1\cup D_2: d_1\perp d_2} = p(\theta)\cdot p(D_1|\theta)\cdot p(D_2|\theta)$$

וקיבלנו את הדרוש.

ii. נרצה למצוא את $p(\theta|D_1, D_2)$. והרי:

$$p(\theta|D_1, D_2)\cdot p(D_1, D_2) = p(\theta, D_1, D_2) = p(\theta)\cdot p(D_1|\theta)\cdot p(D_2|\theta)$$

לכן:

$$p(\theta|D_1, D_2) = \frac{p(\theta)\cdot p(D_1|\theta)\cdot p(D_2|\theta)}{p(D_1, D_2)} \underbrace{=}_{\substack{\forall d_1,d_2\in D_1\cup D_2,\\ d_1\perp d_2}} \frac{p(\theta)\cdot p(D_1|\theta)\cdot p(D_2|\theta)}{p(D_1)\cdot p(D_2)}$$

$$= \frac{p(\theta)\cdot p(D_1|\theta)}{p(D_1)}\cdot\frac{p(D_2|\theta)}{p(D_2)}$$

כלומר נקבל ש:

$$p(\theta|D_1,D_2) \propto \ p(\theta)\cdot p(D_1|\theta)\cdot p(D_2|\theta)$$

והרי אנו יודעים שעבור $\theta \sim \mathcal{N}(\mu,\Sigma)$ מה גם :

$$\forall d_i = (x_i,y_i) \in D_1 \cup D_2 : y_i = \sum_{n=1}^{d} \theta_n h_n(x_i) + \eta \ s.t \ \eta \sim \mathcal{N}(0, I\sigma^2)$$

בפרט עבור $d_i \in D_1$. נגדיר :

$$H_1 = \begin{pmatrix} h^T(x_{1,1}) \\ \dots \\ h^T(x_{1,N_1}) \end{pmatrix}, H_2 = \begin{pmatrix} h^T(x_{2,1}) \\ \dots \\ h^T(x_{2,N_2}) \end{pmatrix}$$

כאשר $\forall j \in \{1,2\} \ : \ N_j = |D_j|, \forall i \in [N_j] : x_{j,i} \in D_j$.

כעת אנו יודעים ש

$$p(\theta|D_1,D_2) = \underbrace{\frac{p(\theta)\cdot p(D_1|\theta)}{p(D_1)}}_{\sim \mathcal{N}(\mu_{\theta|D_1},\Sigma_{\theta|D_1})} \cdot \frac{p(D_2|\theta)}{p(D_2)}$$

ולכן נקבל גם שכלל ההתפלגות גאוסיאנית שהרי מדובר במכפלה בגאוסיאן. נוכל לומר ש :

$$p(\theta|D_1,D_2)\cdot p(\theta) = \frac{p(\theta)\cdot p(D_1|\theta)}{p(D_1)} \cdot \frac{p(\theta)\cdot p(D_2|\theta)}{p(D_2)}$$

$$p(\theta|D_1,D_2) \propto \frac{p(\theta)\cdot p(D_1|\theta)}{p(D_1)} \cdot \frac{p(\theta)\cdot p(D_2|\theta)}{p(D_2)}$$

וכעת נשתמש בידע שלנו על ההתפלגויות $\theta|D_1, \theta|D_2, \theta$ ותכונות האקספוננט :

נקבל שם :

$$\forall i \in [2] : \ p(\theta|D_i) \sim \mathcal{N}(H_i\theta, I\sigma^2)$$

נגדיר :

$$\forall i \in [2] : \ x_i = \begin{bmatrix} x_{i,1} \\ \dots \\ x_{i,N_i} \end{bmatrix}, y_i = \begin{bmatrix} y_{i,1} \\ \dots \\ y_{i,N_i} \end{bmatrix}$$

$$p(\theta|D_1,D_2) \propto \exp\Big[-\frac{1}{2}(y_1 - H_1\theta)^T \frac{1}{\sigma^2} I(y_1 - H_1\theta) - \frac{1}{2}(\theta-\mu)^T \Sigma^{-1}(\theta-\mu) - \frac{1}{2}(y_2 - H_2\theta)^T \frac{1}{\sigma^2} I(y_2 - H_2\theta)\Big]$$

כעת נרצה לגזור את תוכן האקספוננט בכדי לקבל את $\mu_{\theta|D_1,D_2}, \Sigma_{\theta|D_1,D_2}$ על פי טריק הנגזרת.

נגזור את $\{value\ inside\ exponent\}$ קרי :

$$\Delta = -\frac{1}{2}\underbrace{(y_1 - H_1\theta)^T \frac{1}{\sigma^2} I (y_1 - H_1\theta)}_{\Delta_1 = m^T \frac{1}{\sigma^2} I m} - \frac{1}{2}\underbrace{(\theta-\mu)^T \Sigma^{-1} (\theta-\mu)}_{\Delta_2}$$

$$-\frac{1}{2}\underbrace{(y_2 - H_2\theta)^T \frac{1}{\sigma^2} I (y_2 - H_2\theta)}_{\Delta_3}$$

נקבל:

$$\frac{\partial \Delta}{\partial \theta} = \frac{\partial \Delta_1}{\partial \theta} + \frac{\partial \Delta_2}{\partial \theta} + \frac{\partial \Delta_3}{\partial \theta}$$

כעת נחשב:

$$\frac{\partial \Delta_1}{\partial \theta} = \frac{1}{\sigma^2} I \cdot H_1^T (H_1\theta - y_1)$$

$$\frac{\partial \Delta_2}{\partial \theta} = \Sigma^{-1}(\theta - \mu)$$

$$\frac{\partial \Delta_3}{\partial \theta} = \frac{1}{\sigma^2} I \cdot H_2^T (H_2\theta - y_2)$$

כעת:

$$\frac{\partial \Delta}{\partial \theta} = \left(\frac{1}{\sigma^2} I \cdot H_1^T (H_1\theta - y_1) + \Sigma^{-1}(\theta-\mu) + \frac{1}{\sigma^2} I \cdot H_2^T (H_2\theta - y_2)\right)$$

נפתח ונקבל:

$$\frac{\partial \Delta}{\partial \theta} = \frac{1}{\sigma^2} H_1^T H_1 \theta - \frac{1}{\sigma^2} H_1^T y_1 + \frac{1}{\sigma^2} H_2^T H_2 \theta - \frac{1}{\sigma^2} H_2^T y_2 - \Sigma^{-1}\mu + \Sigma^{-1}\theta =$$

$$\frac{\partial \Delta}{\partial \theta} = \left(\frac{1}{\sigma^2} H_1^T H_1 + \frac{1}{\sigma^2} H_2^T H_2 + \Sigma^{-1}\right)\theta - \left[\frac{1}{\sigma^2} H_1^T y_1 + \frac{1}{\sigma^2} H_2^T y_2 - \Sigma^{-1}\mu\right]$$

$$= \left(\overbrace{\frac{1}{\sigma^2} H_1^T H_1 + \frac{1}{\sigma^2} H_2^T H_2 + \Sigma^{-1}}^{\Sigma_{\theta|D_1,D_2}}\right)\left[\theta\right.$$

$$\left. - \underbrace{\left(\frac{1}{\sigma^2} H_1^T y_1 + \frac{1}{\sigma^2} H_2^T y_2 - \Sigma^{-1}\mu\right) \cdot \left(\frac{1}{\sigma^2} H_1^T H_1 + \frac{1}{\sigma^2} H_2^T H_2 + \Sigma^{-1}\right)^{-1}}_{\mu_{\theta|D_1,D_2}}\right]$$

ולכן:

$$p(\theta|D_1,D_2) \sim \mathcal{N}\left(\begin{matrix}\left(\frac{1}{\sigma^2} H_1^T y_1 + \frac{1}{\sigma^2} H_2^T y_2 - \Sigma^{-1}\mu\right) \cdot \left(\frac{1}{\sigma^2} H_1^T H_1 + \frac{1}{\sigma^2} H_2^T H_2 + \Sigma^{-1}\right)^{-1} \\ ,\left(\frac{1}{\sigma^2} H_1^T H_1 + \frac{1}{\sigma^2} H_2^T H_2 + \Sigma^{-1}\right)^{-1}\end{matrix}\right)$$

iii. נניח וביצענו רגרסיה לינארית בייזיאנית כך שיש לנו את:

$$\mu_{\theta,D_1}, \mu_{\theta,D_2}, \Sigma_{\theta|D_1}, \Sigma_{\theta|D_2}$$

נרצה להוכיח שניתן לחשב את $\hat{\theta}^{MMSE} = \mathbb{E}[\theta|D]$ עבור θ נתון בהינתן $\mu_{\theta|D_i}, \Sigma_{\theta|D_i}$: $\forall \in [2]$.

אנו יודעים ש

$$\mu_{\theta|D_1,D_2} = \mathbb{E}[\theta|D] = \hat{\theta}^{MMSE}$$

בנוסף אנו יודעים כי:

$$\frac{\partial \Delta_1}{\partial \theta} = \frac{1}{\sigma^2} I \cdot H_1^T(H_1\theta - y_1) = \underbrace{\frac{1}{\sigma^2}\mathrm{H}_1^{\mathrm{T}}\mathrm{H}_1}_{\Sigma_{\theta|D_1}^{-1} - \Sigma_{\theta}^{-1}}\left(\theta - \underbrace{H_1^{-1}y_1}_{\mu_{\theta|D_1}}\right)$$

$$\frac{\partial \Delta_2}{\partial \theta} = \underbrace{\Sigma^{-1}}_{\Sigma_{\theta}^{-1}}\left(\theta - \underbrace{\mu}_{\mu_{\theta}}\right)$$

$$\frac{\partial \Delta_3}{\partial \theta} = \frac{1}{\sigma^2} I \cdot H_2^T(H_2\theta - y_2) = \underbrace{\frac{1}{\sigma^2}\mathrm{H}_2^{\mathrm{T}}\mathrm{H}_2}_{\Sigma_{\theta|D_1}^{-1} - \Sigma_{\theta}^{-1}}\left(\theta - \underbrace{H_2^{-1}y_2}_{\mu_{\theta|D_2}}\right)$$

כעת נבחין כי:

$$\left(\frac{1}{\sigma^2}\mathrm{H}_1^{\mathrm{T}}\mathrm{H}_1 + \Sigma^{-1} + \frac{1}{\sigma^2}\mathrm{H}_2^{\mathrm{T}}\mathrm{H}_2\right) = \Sigma_{\theta|D_1}^{-1} + \Sigma_{\theta|D_2}^{-1} - \Sigma_{\theta}^{-1} = \Sigma_{\theta|D_1,D_2}^{-1}$$

$$\underbrace{\frac{1}{\sigma^2}H_1^T y_1}_{\frac{1}{\sigma^2}\mathrm{H}_1^{\mathrm{T}}\mathrm{H}_1 \cdot H_1^{-1}y_1} + \Sigma^{-1}\mu + \underbrace{\frac{1}{\sigma^2}H_2^T y_2}_{\frac{1}{\sigma^2}\mathrm{H}_2^{\mathrm{T}}\mathrm{H}_2 \cdot H_2^{-1}y_2} = \Sigma_{\theta|D_1}^{-1} \cdot \mu_{\theta|D_1} - \Sigma_{\theta}^{-1}\mu_{\theta} + \Sigma_{\theta|D_2}^{-1} \cdot \mu_{\theta|D_2}$$

והרי

$$\hat{\theta}^{MMSE} = \mathbb{E}[\theta|D] = \left(\underbrace{\frac{1}{\sigma^2}H_1^T y_1 + \frac{1}{\sigma^2}H_2^T y_2 + \Sigma^{-1}\mu}_{\Sigma_{\theta|D_1}^{-1} \cdot \mu_{\theta|D_1} + \Sigma_{\theta}^{-1}\mu_{\theta} + \Sigma_{\theta|D_2}^{-1} \cdot \mu_{\theta|D_2}}\right) \cdot \underbrace{\left(\frac{1}{\sigma^2}H_1^T H_1 + \frac{1}{\sigma^2}H_2^T H_2 + \Sigma^{-1}\right)^{-1}}_{\left(\Sigma_{\theta|D_1}^{-1} - \Sigma_{\theta}^{-1} + \Sigma_{\theta|D_2}^{-1}\right)^{-1}}$$

ולכן:

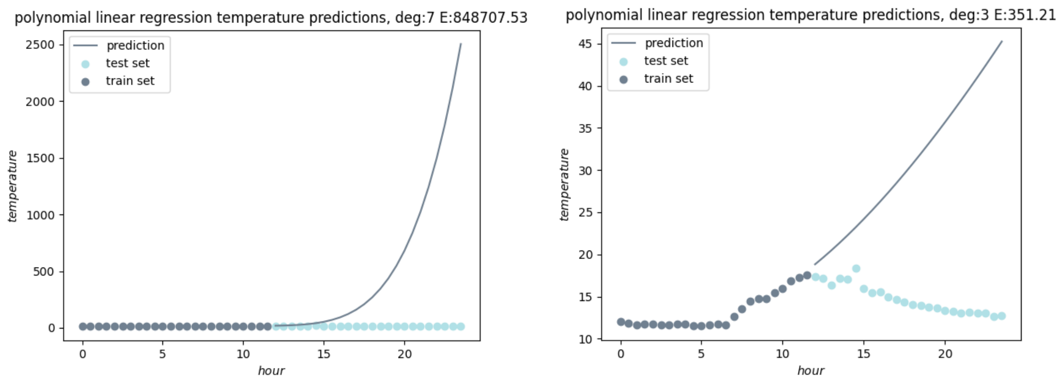
$$\hat{\theta}^{MMSE} = \left(\Sigma_{\theta|D_1}^{-1} \cdot \mu_{\theta|D_1} - \Sigma_{\theta}^{-1}\mu_{\theta} + \Sigma_{\theta|D_2}^{-1} \cdot \mu_{\theta|D_2}\right) \cdot \left(\Sigma_{\theta|D_1}^{-1} + \Sigma_{\theta|D_2}^{-1} - \Sigma_{\theta}^{-1}\right)^{-1}$$

2. <u>חלק פרקטי</u>

2.1. היינו צריכים למדל רגרסיה לינארית קלאסית. זה נעשה במחלקה הניתנה לנו.

2.2. היינו צריכים למדל רגרסיה לינארית בייזיאנית, זה נעשה במחלקה הניתנה לנו.

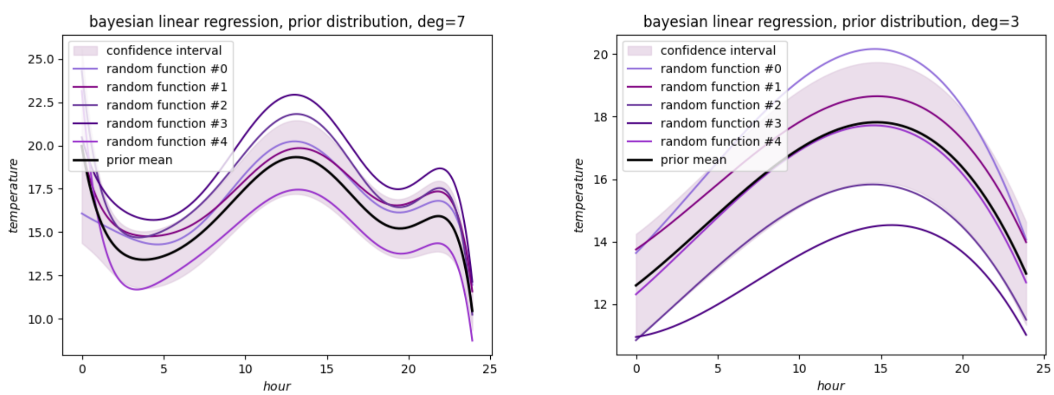
2.3. באימון המודל הקלאסי קיבלנו את התוצאות הבאות (ביחס לdeg שניתן לנו):



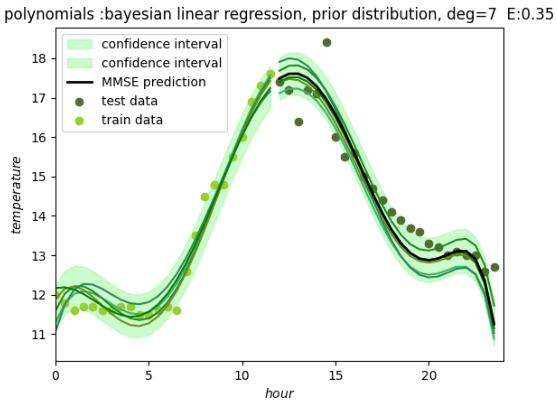
בכותרת הגרף מצורפים ערכי ה MSE עבור כל מודל כתלות בdeg.

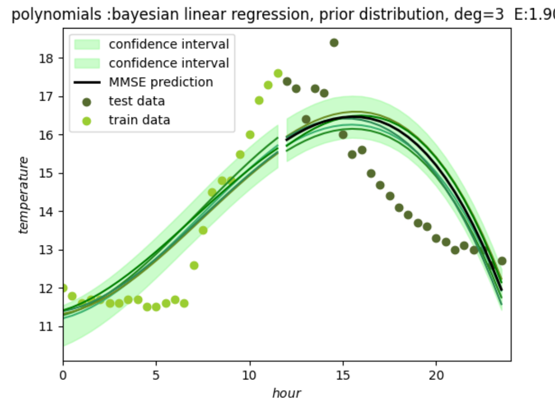
2.4 אימנו $prior$.

2.5 נייצג את $mean\ functions$ שמוגדרות על ידי ה$prior$ ביחד עם רווח הסמך ובנוסף 5 פונקציות בעזרת ה$prior$שלמדנו, ביחס לdeg:



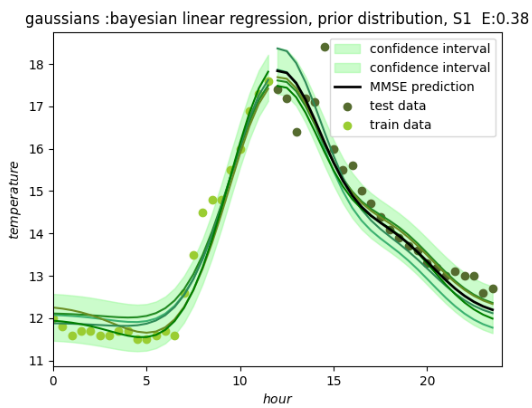
2.6 נתאים את המודל ל$test$ ונקבל את ה$posterior$. נייצג את ה$true\ test$ ביחד עם הפרדיקציה עם הרווח הסמך מסביב אליה, ובנוסף 5 פונקציות נוספות שנלמדו בעזרת ה $posterior$, ביחס לdeg:

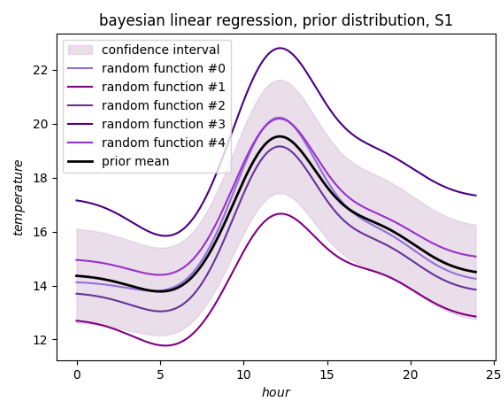


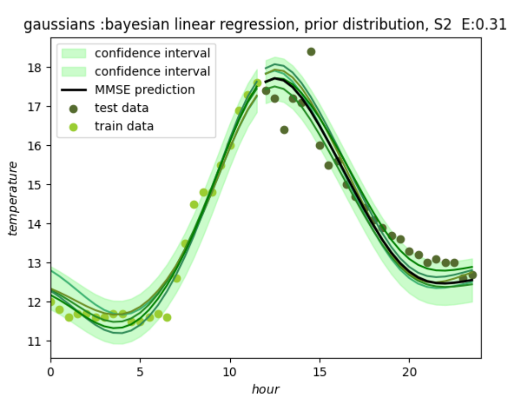


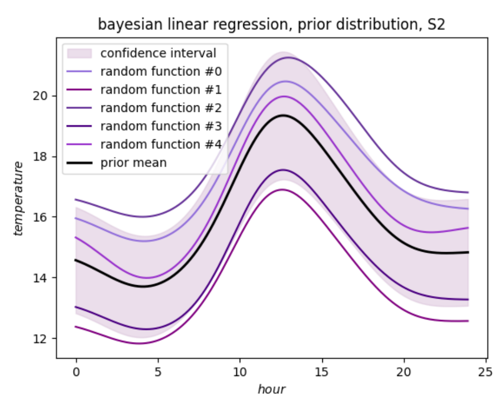
הMSE של ה $MMSE$ מצורפת בכותרת הplot.

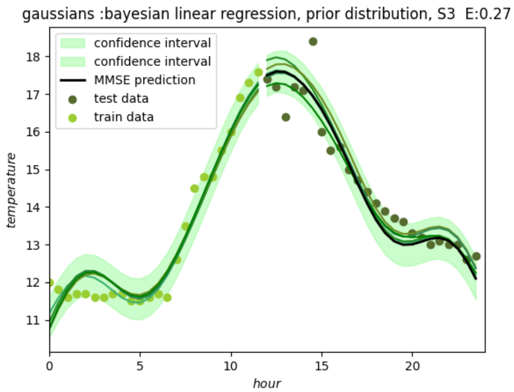
2.7 נבצע את סעיפים 4-6 על פונקציות בסיס גאוסיאניות :

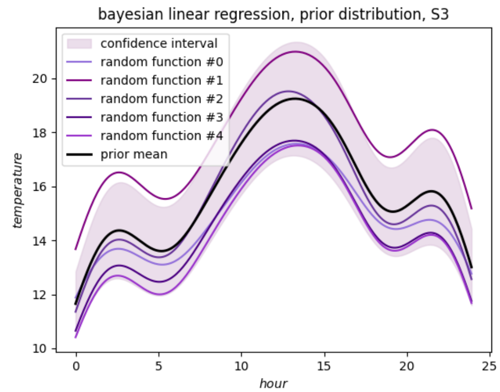




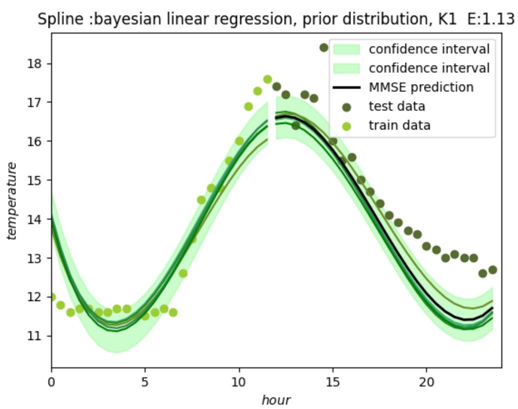


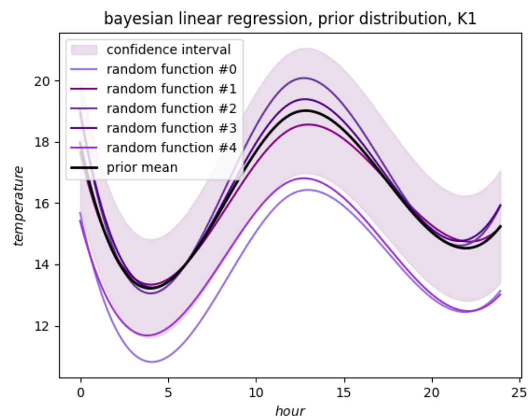


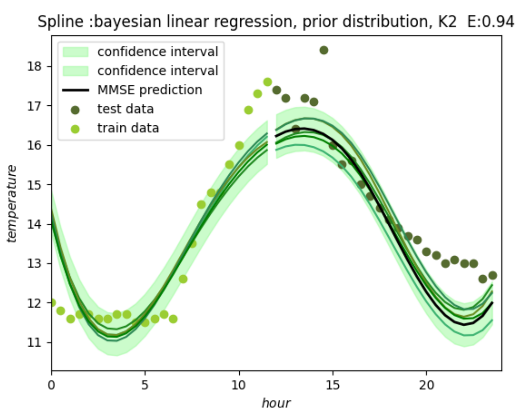


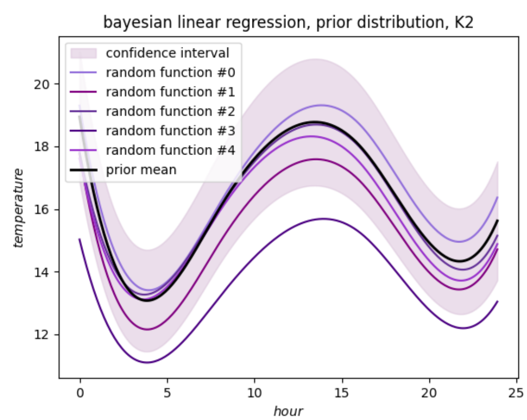


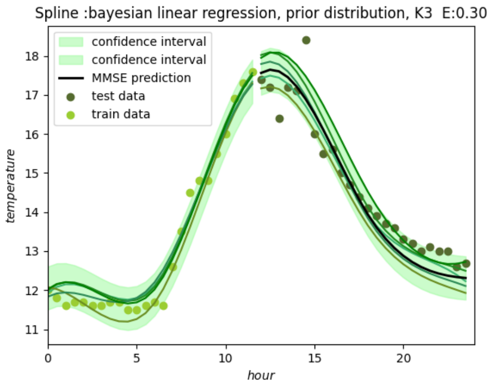
2.8 נבצע את סעיפים 4-6 על פונקציות בסיס $Cubib\ Splines$:

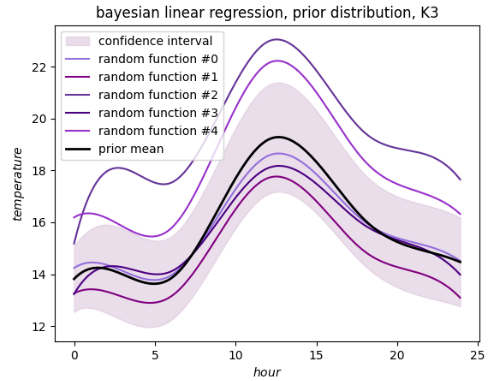












2.9 נתבונן על הMSE של הפתרונות שלנו :

```
Average squared error with LR and d=3 is 351.21
Average squared error with LR and d=7 is 848707.53
Average squared error with BLR and deg=3  is 1.90
Average squared error with BLR and deg=7  is 0.35
Average squared error with BLR and S1  is 0.38
Average squared error with BLR and S2  is 0.31
Average squared error with BLR and S3  is 0.27
Average squared error with BLR and K1  is 1.13
Average squared error with BLR and K2  is 0.94
Average squared error with BLR and K3  is 0.30
```