COMS 4771 HW 3

Griffin Klett

TOTAL POINTS

108 / 110

QUESTION 1

Privacy in data 20 pts

1.1 (i)(a) 6 / 6

- ✓ **+ 2 pts** Correctly sets up problem
- ✓ **+ 1 pts** Shows reasonable derivation
- ✓ **+ 3 pts** Arrives at correct result
- **+ 0 pts** Incorrect
- **+ 0 pts** Missing
- **- 1 pts** Minor error in setup (a possible typo)
- **- 1 pts** Minor error in final result (a possible typo)

1.2 (i)(b) 2 / 2

- ✓ **- 0 pts** $$\frac{k+1}{2}$$
- **- 1 pts** Minor error
- **- 2 pts** Incorrect
- **- 2 pts** Missing

1.3 (ii)(a) 6 / 6

- ✓ **+ 2 pts** Correct setup
- ✓ **+ 2 pts** Shows complete derivation
- ✓ **+ 2 pts** Arrives at correct final result
- **+ 1 pts** Incorrect with reasonable work
- **+ 0 pts** Missing
- **- 1 pts** Minor error on pdf (possible typo)

1.4 (ii)(b) 6 / 6

- ✓ **+ 3 pts** Correct proof of the tail bound
- ✓ **+ 3 pts** Correct application of the tail bound to the problem
- **+ 1 pts** Partially correct/complete proof of the tail bound
- **+ 1 pts** Partially correct/complete application of the tail bound
- **- 1 pts** Minor error in proof of tail bound
- **- 1 pts** Minor error in application of tail bound
- **+ 0 pts** Missing/Not Attempted

QUESTION 2

Non-parametric Regression via Bayesian Modelling 40 pts

2.1 (i) 3 / 3

- ✓ **- 0 pts** Correct
- **- 1 pts** Minor error
- **- 2 pts** Reasonable attempt / correct but no work
- **- 3 pts** Missing / incomplete attempt

2.2 (ii) 4 / 4

- ✓ **- 0 pts** Correct
- **- 1 pts** Correct with minor error
- **- 3 pts** Reasonable attempt
- **- 4 pts** Missing / incomplete attempt

2.3 (iii) 4 / 4

- ✓ **- 0 pts** Correct
- **- 1 pts** Correct with minor error
- **- 3 pts** Reasonable attempt
- **- 4 pts** Missing / incomplete attempt

2.4 (iv) 2 / 2

- ✓ **+ 2 pts** Correct answer through marginal
- **+ 0 pts** Incorrect or Missing
- **+ 1 pts** Almost correct (small typo)

2.5 (v) 4 / 4

- ✓ **+ 4 pts** +1 Plots, +1 smoothness comment, +1 what happens when all ones (constant), +1 some valid explanation about mu AND sigma role OR extra experiment which indicates this trend
- **+ 3 pts** 3/4 of above
- **+ 2 pts** 2/4 of above

+ **1 pts** 1/4 of above
+ **0 pts** Incorrect or missing

2.6 (vi) 4 / 4

✓ + **4 pts** **Correct: BOTH 1 and 2**
+ **3 pts** Either:
1. Only smoothness description AND rationalization but without plots
2. Only plots with description but without reasonable justification for smoothness
+ **1 pts** Plots appear right but incorrect/missing description
+ **0 pts** Incorrect/Missing

2.7 (vii) 4 / 4

✓ - **0 pts** **Correct**
- **2 pts** No explanation about settings of $$\Sigma$$ and $$\mu$$
- **2 pts** No plot/Incorrect plot
- **4 pts** Incorrect/Missing

2.8 (viii) 3 / 3

✓ - **0 pts** **Correct**
- **1 pts** Minor issue
- **3 pts** Incorrect/Missing

2.9 (ix) 4 / 4

✓ - **0 pts** **Correct**
- **2 pts** Incorrect/Missing written part for this question.
- **2 pts** Incorrect/Missing plots(Curves didn't pass through training data or missing training data on plots)
- **4 pts** Incorrect/Missing

2.10 (x) 4 / 4

✓ + **4 pts** **(+2)4 plots through all training data points with changed period OR plots through some training data using pseudo-inverse (WITH written acknowledgement of issue)**
(+1) overlay of points on graph as proof
(+1) correct written reflection on the graph

+ **3 pts** Any three points scored in above rubric
+ **2 pts** Any two points scored in above rubric
+ **1 pts** Single point scored in above rubric
+ **0 pts** Incorrect/missing

2.11 (xi) 2 / 2

✓ - **0 pts** **Correct**
- **2 pts** Incorrect/Missing

2.12 (xii) 0 / 2

- **0 pts** Correct
- **1 pts** Incorrect/Missing one plot

✓ - **2 pts** **Incorrect/Missing**

QUESTION 3

3 Regression Competition 50 / 50

✓ - **0 pts** **Comprehensive PDF**
- **5 pts** Slightly incomplete/vague explanation [why did you pick a model?/what else did you try?]
- **12.5 pts** Incomplete/Vague explanation [what models did you pick, what experiments did you run?]
- **20 pts** Minimum PDF submisison credit
- **25 pts** No PDF submission found
- **0 pts** Click here to replace this description.

Griffin Klett

Gk2591

COMS 4771 SP21 HW3

3/22/21

1. Privacy in data

i. a. Privacy Guarantee of RR

If we want RR to be $(\epsilon, 0)$ differentially private then we can write our equation given the definition as:

$$e^{-\epsilon} \leq \frac{P[RR(x) = r]}{P[RR(y) = r]} \leq e^{\epsilon}$$

For any two adjacent databases $x, y \epsilon \mathbb{N}^{|\mathcal{X}|}$

Where r is $r \epsilon \{Yes, No\}$

And our RR is a randomized algorithm which maps to True/False:

$$\mathbb{N}^{|\mathcal{X}|} \rightarrow \{ Yes, No \}$$

Since r can take only two values and we are calculating for two adjacent DBs, there are four probabilities to calculate: $P[RR(x) = Yes], P[RR(x) = no]\ for\ x\ and\ y$

…

When sensitive_question(x) = True is given by two cases:

RR(x) returns true when $m < k$ and when $m \geq k\ \&\&\ heads$

The probabilities we need here are based on m and k. Since k is a real number in $[0,1]$ and m is uniformly sampled from $[0,1]$ we can write:

$$P(m < k) = k$$

And:

$$P(m \geq k) = 1 - k$$

Back to sensitive_question(x) = True will return Yes:

$$P[RR(x) = Yes \mid sensitiveQuestion = True] = P(m < k) + P(m \geq k, heads)$$

$$= k + (1 - k) * \frac{1}{2}$$

$$= \frac{1 + k}{2}$$

And sensitive_question(x) = True will return No in two cases also:

$$P[RR(x) = No \mid sensitiveQuestion = True] = P(m \geq k, tails)$$
$$= \frac{1-k}{2}$$

Senstive question(x) = False will return Yes:

$$P[RR(x) = Yes \mid sensitiveQuestion = False] = P(m \geq k, heads)$$
$$= \frac{1-k}{2}$$

Senstive question(x) = False will return No:

$$P[RR(x) = No \mid sensitiveQuestion = False] = P(m < k) + P(m \geq k, tails)$$
$$= k + \frac{1-k}{2}$$
$$= \frac{1+k}{2}$$

Now given these probabilities we can apply them to the original equation given adjacent $x, y \in \mathbb{N}^{|X|}$

Case 1: sensitiveQuestion(x) = True, sensitiveQuestion(y) = True (same value)

$$\frac{P[RR(x) = Yes]}{P[RR(y) = Yes]} = \frac{P[RR(x) = Yes) |\ senstiveQuestion(x) = Trye]}{P[RR(y) = Yes|sensitiveQuestion(y) = True]} = \frac{\frac{1+k}{2}}{\frac{1+k}{2}} = 1$$

$$\frac{P[RR(x) = No]}{P[RR(y) = No]} = \frac{P[RR(x) = No) |\ senstiveQuestion(x) = Trye]}{P[RR(y) = No|sensitiveQuestion(y) = True]} = \frac{\frac{1-k}{2}}{\frac{1-k}{2}} = 1$$

The same output (1) applies when both sensitiveQuestion(x/y) = False. Only showing true case for brevity

Substuting the value of 1 into the original equation leaves us with:

$$e^{-\epsilon} \leq 1 \leq e^{\epsilon}$$

Case 2: SensitiveQuestion(x) = True, SensitiveQuestion(y) = False (Different signs)

$$\frac{P[RR(x) = Yes]}{P[RR(y) = Yes]} = \frac{P[RR(x) = Yes)|\ senstiveQuestion(x) = True]}{P[RR(y) = Yes|sensitiveQuestion(y) = False]} = \frac{\frac{1+k}{2}}{\frac{1-k}{2}} = \frac{1+k}{1-k}$$

$$\frac{P[RR(x) = No]}{P[RR(y) = No]} = \frac{P[RR(x) = No)|\ senstiveQuestion(x) = True]}{P[RR(y) = No|sensitiveQuestion(y) = False]} = \frac{\frac{1-k}{2}}{\frac{1+k}{2}} = \frac{1-k}{1+k}$$

The same applies to when SensitiveQuestion(x) = False and SensitiveQuestion(y) = True. Only showing one for brevity

Substituting these values into original equation leaves us with two additional cases:

$$e^{-\epsilon} \leq \frac{1+k}{1-k} \leq e^{\epsilon}$$

And:

$$e^{-\epsilon} \leq \frac{1-k}{1+k} \leq e^{\epsilon}$$

In order to meet the condition:

$$e^{-\epsilon} \leq 1 \leq e^{\epsilon}$$

We know $\epsilon \geq 0$

Now rearranging the first case for k:

$$e^{-\epsilon} \leq \frac{1+k}{1-k} \leq e^{\epsilon}$$

$$e^{-\epsilon} \leq \frac{2}{1-k} - 1 \leq e^{\epsilon}$$

$$e^{-\epsilon} + 1 \leq \frac{2}{1-k} \leq e^{\epsilon} + 1$$

$$\frac{1}{e^{-\epsilon} + 1} \leq \frac{1-k}{2} \leq \frac{1}{e^{\epsilon} + 1}$$

$$\frac{2}{e^{-\epsilon} + 1} \leq 1 - k \leq \frac{2}{e^{\epsilon} + 1}$$

$$1-\frac{2}{e^{-\epsilon}+1}\leq k\leq 1-\frac{2}{e^{\epsilon}+1}$$

$$\frac{e^{-\epsilon}-1}{e^{-\epsilon}+1}\leq k\leq \frac{e^{\epsilon}-1}{e^{\epsilon}+1}$$

The left side, even with a minimum value $\epsilon = 0$ returns:

$$\frac{1-1}{2}=0$$

so we can rewrite as:

$$0\leq k\leq \frac{e^{\epsilon}-1}{e^{\epsilon}+1}$$

Which gives the values of k that would make $RR(\epsilon, 0)$ differentially private

1.i.b. Accuracy Guarantee of RR

$$\begin{aligned}P[sensitiveQuestion(x) = RR(x)]\\ &= P[sensitiveQuestion(x) = True, RR(x) = Yes]\\ &+ P[sensitiveQuestion(x) = False, RR(x) = No]\end{aligned}$$

Given our RR(x) function, if $m < k$, then $P[sensitiveQuestion(x) = RR(x)] = 1$

If $m \geq k, then\ P[SensitiveQuestion(x) = RR(x) = \frac{1}{2}$

Apply conditional probabilities:

$$P(m < k) = k$$

$$P(m \geq k) = 1 - k$$

$$P[sensitiveQuestion(x) = RR(x)] = k * 1 + (1-k) * \frac{1}{2} = \frac{1+k}{2}$$

1.ii.a Privacy Guarantee of RR:

Let $\delta = 0$

Given any function $f\colon \mathbb{N}^{|\chi|} \to \mathbb{R}, define$:

$$\mathcal{M}_L(x, f(*), \epsilon) = f(x) + Y$$

1.1 (i)(a) **6 / 6**

- ✓ **+ 2 pts** **Correctly sets up problem**
- ✓ **+ 1 pts** **Shows reasonable derivation**
- ✓ **+ 3 pts** **Arrives at correct result**
- **+ 0 pts** Incorrect
- **+ 0 pts** Missing
- **- 1 pts** Minor error in setup (a possible typo)
- **- 1 pts** Minor error in final result (a possible typo)

$$1 - \frac{2}{e^{-\epsilon} + 1} \leq k \leq 1 - \frac{2}{e^{\epsilon} + 1}$$

$$\frac{e^{-\epsilon} - 1}{e^{-\epsilon} + 1} \leq k \leq \frac{e^{\epsilon} - 1}{e^{\epsilon} + 1}$$

The left side, even with a minimum value $\epsilon = 0$ returns:

$$\frac{1 - 1}{2} = 0$$

so we can rewrite as:

$$0 \leq k \leq \frac{e^{\epsilon} - 1}{e^{\epsilon} + 1}$$

Which gives the values of k that would make $RR(\epsilon, 0)$ differentially private

1.i.b. Accuracy Guarantee of RR

$$\begin{aligned} P[sensitiveQuestion(x) = RR(x)] \\ &= P[sensitiveQuestion(x) = True, RR(x) = Yes] \\ &+ P[sensitiveQuestion(x) = False, RR(x) = No] \end{aligned}$$

Given our RR(x) function, if $m < k$, then $P[sensitiveQuestion(x) = RR(x)] = 1$

If $m \geq k, then\ P[SensitiveQuestion(x) = RR(x) = \frac{1}{2}$

Apply conditional probabilities:

$$P(m < k) = k$$

$$P(m \geq k) = 1 - k$$

$$P[sensitiveQuestion(x) = RR(x)] = k * 1 + (1 - k) * \frac{1}{2} = \frac{1 + k}{2}$$

1.ii.a Privacy Guarantee of RR:

Let $\delta = 0$

Given any function $f\colon \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}, define$:

$$\mathcal{M}_L(x, f(*), \epsilon) = f(x) + Y$$

1.2 (i)(b) **2 / 2**

✓ **- 0 pts** **$$\frac{k+1}{2}$$**

- 1 pts Minor error

- 2 pts Incorrect

- 2 pts Missing

$$1 - \frac{2}{e^{-\epsilon} + 1} \leq k \leq 1 - \frac{2}{e^{\epsilon} + 1}$$

$$\frac{e^{-\epsilon} - 1}{e^{-\epsilon} + 1} \leq k \leq \frac{e^{\epsilon} - 1}{e^{\epsilon} + 1}$$

The left side, even with a minimum value $\epsilon = 0$ returns:

$$\frac{1 - 1}{2} = 0$$

so we can rewrite as:

$$0 \leq k \leq \frac{e^{\epsilon} - 1}{e^{\epsilon} + 1}$$

Which gives the values of k that would make $RR(\epsilon, 0)$ differentially private

1.i.b. Accuracy Guarantee of RR

$$\begin{aligned} P[sensitiveQuestion(x) = RR(x)] \\ &= P[sensitiveQuestion(x) = True, RR(x) = Yes] \\ &+ P[sensitiveQuestion(x) = False, RR(x) = No] \end{aligned}$$

Given our RR(x) function, if $m < k$, then $P[sensitiveQuestion(x) = RR(x)] = 1$

If $m \geq k, then\ P[SensitiveQuestion(x) = RR(x) = \frac{1}{2}$

Apply conditional probabilities:

$$P(m < k) = k$$

$$P(m \geq k) = 1 - k$$

$$P[sensitiveQuestion(x) = RR(x)] = k * 1 + (1 - k) * \frac{1}{2} = \frac{1 + k}{2}$$

1.ii.a Privacy Guarantee of RR:

Let $\delta = 0$

Given any function $f: \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}, define$:

$$\mathcal{M}_L(x, f(*), \epsilon) = f(x) + Y$$

Where Y is a random variable drawn from $L(y; \frac{\Delta f}{\epsilon})$ where

$$L(y; b) = \frac{1}{2b} \exp\left(-\frac{|y|}{b}\right)$$

So:

$$\mathcal{M}_L(x, f(*), \epsilon) = f(x) + Y$$

Where:

$$L\left(y; \frac{\Delta f}{\epsilon}\right) = \frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon|y|}{\Delta f}\right)$$

To find the privacy, we will do the equivalent of the discontinuous case given by:

$$\ln\left(\frac{P[\mathcal{M}(x) = r]}{P[\mathcal{M}(y) = r]}\right)$$

By replacing the probabilities with densities

$$\frac{p_M(x)}{P_M(y)}$$

We need to evaluate the densities at an arbitrary point $h \epsilon \mathbb{R}$ in order to get a ratio; from this we can tell the privacy

$$p_{M(x)}(h) = p_{M(x)}(f(x) + Y = h) = p_Y(Y = h - f(x))$$

Which we can rewrite using L as:

$$p_{M(x)}(h) = \frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon|z - f(x)|}{\Delta f}\right)$$

And similarly for $p_{M(y)}$:

$$p_{M(y)}(h) = \frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon|z - f(y)|}{\Delta f}\right)$$

Rewriting the ratio now and simplifying:

$$\frac{p_{M(x)}(h)}{p_{M(y)}(h)} = \frac{\frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon|z - f(x)|}{\Delta f}\right)}{\frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon|z - f(y)|}{\Delta f}\right)} = \exp\left(\frac{(\epsilon|z - f(y)| - |z - f(x)|)}{\Delta f}\right)$$

Using triangle equality on quantity inside exponent yields:

$$\frac{p_{M(x)}(h)}{p_{M(y)}(h)} = \frac{\frac{\epsilon}{2\Delta f}\exp\left(-\frac{\epsilon|z-f(x)|}{\Delta f}\right)}{\frac{\epsilon}{2\Delta f}\exp\left(-\frac{\epsilon|z-f(y)|}{\Delta f}\right)} \leq \exp\left(\frac{(\epsilon|f(x)-f(y)|)}{\Delta f}\right)$$

Since $(\max\{|f(x)-f(y)|\} = \Delta f$

$$\frac{p_{M(x)}(h)}{p_{M(y)}(h)} = \exp\left(\frac{(\epsilon * \Delta f)}{\Delta f}\right) = \exp(\epsilon)$$

Thus we can achieve $(\epsilon, 0)$ differential privacy

1.ii.b Accuracy Guarantee of A

Note that we defined $\mathcal{M}_L(x, f(*), \epsilon)$:

$$\mathcal{M}_L(x, f(*), \epsilon) = f(x) + Y$$

Rearranging yields:

$$Y = f(x) - \mathcal{M}_L(x, f(*), \epsilon)$$

So we can rewrite the original equation as:

$$P[|f(x) - \mathcal{M}_L(x, f(*), \epsilon)| \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right) = P\left[|Y| \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right)\right] = p$$

The idea is to first prove

$$P[Y \geq bt] = \exp(-t)$$

Then substitute that into:

$$P[Y \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right)]$$

Where $b = \frac{\Delta f}{\epsilon}$ and $t = \ln\left(\frac{1}{p}\right)]$

…

Looking at $P[Y \geq bt] = \exp(-t)$ we see that $L(y; b)$ is symmetric and also that $b > 0$ since we want the output to be be a positive (density) and that $t > 0$ otherwise $P[Y \geq bt]$ would always equal 1. so we can rewrite as:

$$P[|Y| \geq bt] = P[Y \geq bt] + P[Y \leq -bt] = 2P[Y \geq bt]$$

Substitute in the value of L(y;b) and set up integral:

1.3 (ii)(a) **6 / 6**

- ✓ **+ 2 pts** Correct setup
- ✓ **+ 2 pts** Shows complete derivation
- ✓ **+ 2 pts** Arrives at correct final result
- **+ 1 pts** Incorrect with reasonable work
- **+ 0 pts** Missing
- **- 1 pts** Minor error on pdf (possible typo)

Using triangle equality on quantity inside exponent yields:

$$\frac{p_{M(x)}(h)}{p_{M(y)}(h)} = \frac{\frac{\epsilon}{2\Delta f}\exp\left(-\frac{\epsilon|z-f(x)|}{\Delta f}\right)}{\frac{\epsilon}{2\Delta f}\exp\left(-\frac{\epsilon|z-f(y)|}{\Delta f}\right)} \leq \exp\left(\frac{(\epsilon|f(x)-f(y)|)}{\Delta f}\right)$$

Since $(\max\{|f(x)-f(y)|\} = \Delta f$

$$\frac{p_{M(x)}(h)}{p_{M(y)}(h)} = \exp\left(\frac{(\epsilon * \Delta f)}{\Delta f}\right) = \exp(\epsilon)$$

Thus we can achieve $(\epsilon, 0)$ differential privacy

1.ii.b Accuracy Guarantee of A

Note that we defined $\mathcal{M}_L(x, f(*), \epsilon)$:

$$\mathcal{M}_L(x, f(*), \epsilon) = f(x) + Y$$

Rearranging yields:

$$Y = f(x) - \mathcal{M}_L(x, f(*), \epsilon)$$

So we can rewrite the original equation as:

$$P[|f(x) - \mathcal{M}_L(x, f(*), \epsilon)| \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right) = P\left[|Y| \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right)\right] = p$$

The idea is to first prove

$$P[Y \geq bt] = \exp(-t)$$

Then substitute that into:

$$P[Y \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right)]$$

Where $b = \frac{\Delta f}{\epsilon}$ and $t = \ln\left(\frac{1}{p}\right)]$

…

Looking at $P[Y \geq bt] = \exp(-t)$ we see that $L(y; b)$ is symmetric and also that $b > 0$ since we want the output to be be a positive (density) and that $t > 0$ otherwise $P[Y \geq bt]$ would always equal 1. so we can rewrite as:

$$P[|Y| \geq bt] = P[Y \geq bt] + P[Y \leq -bt] = 2P[Y \geq bt]$$

Substitute in the value of L(y;b) and set up integral:

$$P[|Y| \geq bt] = \frac{1}{b}\int_{bt}^{\infty} \exp\,(-\frac{y}{b}dy)$$

(y is only positive in this integral because of the above condition that b and t are also already positive)

Integrate to get the desired output:

$$P[Y \geq bt] = \exp\,(-t)$$

Now we can evaluate the original equation while substituting $b = \frac{\Delta f}{\epsilon}$ and $t = \ln\left(\frac{1}{p}\right)$]

$$P[|f(x) - \mathcal{M}_L(x, f(*), \epsilon)| \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right) = P\left[|Y| \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right)\right] = \exp\,(-\ln\left(\frac{1}{p}\right))$$

$$[|f(x) - \mathcal{M}_L(x, f(*), \epsilon)| \geq \left(\frac{\Delta f}{\epsilon}\right)\ln\left(\frac{1}{p}\right) = \frac{1}{\exp\left(\ln\left(\frac{1}{p}\right)\right)} = p$$

1.4 (ii)(b) **6 / 6**

- ✓ **+ 3 pts** **Correct proof of the tail bound**
- ✓ **+ 3 pts** **Correct application of the tail bound to the problem**
- **+ 1 pts** Partially correct/complete proof of the tail bound
- **+ 1 pts** Partially correct/complete application of the tail bound
- **- 1 pts** Minor error in proof of tail bound
- **- 1 pts** Minor error in application of tail bound
- **+ 0 pts** Missing/Not Attempted

2. Non-parametric Regression via Bayesian Modeling

Reference/Citation: Please note, I followed the proof steps from the lecture notes here for parts ii-iv: http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html

2.i. Derive the marginal distribution of x_i

Starting with the equation from part 2.iii. (proved below):

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}Q(x_1,x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right) * \frac{1}{(2\pi)^{\frac{d}{2}}|A|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

$$f_1(x_1) = \int f(x_1,x_2)dx_2$$

$$= \int \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right) * \frac{1}{(2\pi)^{\frac{d}{2}}|A|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

Pull out x_1 terms:

$$f_1(x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right) * \int \frac{1}{(2\pi)^{\frac{d}{2}}|A|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)dx_2$$

Since the integral is a probability density function over x_2, the integral evaluates to 1, leaving us with:

$$f_1(x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right) * 1$$

2.ii.

starting from the joint density of x:

$$f(\vec{x}) = f(x_1,x_2) = \frac{1}{(2\pi)^{\frac{\pi}{2}}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\vec{x}-\mu)^T\Sigma^{-1}(\vec{x}-\mu)\right] = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}Q(x_1,x_2)\right)$$

Where Q is defined as:

$$Q(x_1,x_2) = (\vec{x}-\mu)^T\Sigma^{-1}(\vec{x}-\mu) = [(x_1-\mu_1)^T,(x_2-\mu_2)^T]\begin{bmatrix}\Sigma^{11} & \Sigma^{12}\\ \Sigma^{21} & \Sigma^{22}\end{bmatrix}\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix} =$$
$$(x_1-\mu_1)^T\Sigma^{11}(x_1-\mu_1) + 2(x_1-\mu_1)^T\Sigma^{12}(x_2-\mu_2) + (x_2-\mu_2)^T\Sigma^{22}(x_2-\mu_2)$$

2.1 (i) **3 / 3**

- ✓ **- 0 pts** Correct
- **- 1 pts** Minor error
- **- 2 pts** Reasonable attempt / correct but no work
- **- 3 pts** Missing / incomplete attempt

2. Non-parametric Regression via Bayesian Modeling

Reference/Citation: Please note, I followed the proof steps from the lecture notes here for parts ii-iv: http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html

2.i. Derive the marginal distribution of x_i

Starting with the equation from part 2.iii. (proved below):

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}Q(x_1,x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right.$$
$$* \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

$$f_1(x_1) = \int f(x_1,x_2)dx_2$$

$$= \int \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right. * \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

Pull out x_1 terms:

$$f_1(x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right.$$
$$* \int \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)dx_2$$

Since the integral is a probability density function over x_2, the integral evaluates to 1, leaving us with:

$$f_1(x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right. * 1$$

2.ii.

starting from the joint density of x:

$$f(\vec{x}) = f(x_1,x_2) = \frac{1}{(2\pi)^{\frac{\pi}{2}}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\vec{x}-\mu)^T\Sigma^{-1}(\vec{x}-\mu)\right] = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}Q(x_1,x_2)\right)$$

Where Q is defined as:

$$Q(x_1,x_2) = (\vec{x}-\mu)^T\Sigma^{-1}(\vec{x}-\mu) = [(x_1-\mu_1)^T,(x_2-\mu_2)^T]\begin{bmatrix}\Sigma^{11} & \Sigma^{12}\\ \Sigma^{21} & \Sigma^{22}\end{bmatrix}\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix} =$$
$$(x_1-\mu_1)^T\Sigma^{11}(x_1-\mu_1) + 2(x_1-\mu_1)^T\Sigma^{12}(x_2-\mu_2) + (x_2-\mu_2)^T\Sigma^{22}(x_2-\mu_2)$$

Where we assume

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$$

Now substitute the three facts into the equation for $Q(x_1, x_2)$:

$$Q(x_1, x_2) = (\overrightarrow{x_1} - \mu_1)^T [\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}](x_1 - \mu_1)$$

$$-2(x_1 - \mu_1)^T [\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2 - \mu_2)$$

$$+(x_2 - \mu_2)^T [(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2 - \mu_2)$$

And expanding out:

$$Q(x_1, x_2) = (x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) +$$

$$(x_1 - \mu_1)^T [\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}](x_1 - \mu_1)$$

$$-2(x_1 - \mu_1)^T [\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2 - \mu_2)$$

$$+(x_2 - \mu_2)^T [(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2 - \mu_2)$$

And rearranging:

$$Q(x_1, x_2) = (x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) +$$
$$[(x_2 - \mu_2) - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1)]^T (\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}[(x_2 - \mu_2) - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1)]$$

Now we define A, b

$$A = \Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}$$
$$b = \mu_2 - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1)$$

And re-write $Q(x_1, x_2)$:

$$Q(x_1, x_2) = (x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) + (x_2 - b)^T A^{-1}(x_2 - b)$$

Plug this value of $Q(x_1, x_2)$ into the original joint distribution eqn to get the desired result:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1, x_2)\right)$$

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1) + (x_2-b)^T A^{-1}(x_2-b)\right)$$

iii.

substitute in the given fact:

$$|\Sigma| = |\Sigma_{11}|\left|\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right|$$

into our equation for joint density:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right)$$

Substitute value of A:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right)$$

And split the $Q(x_1,x_2)$ into $Q(x_1)+Q(x_1,x_2)$:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right.$$
$$* \frac{1}{(2\pi)^{\frac{d}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

$$f(\vec{x}) = \aleph(x_1,\mu_1,\Sigma_{11}) * \aleph(x_2,b,A)$$

iv.

to find the conditional distribution of x_2 given x_1, divide the joint distribution by the marginal distriubiton for x_1 (from part 1):

$$f_1(x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right.$$

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right. * \frac{1}{(2\pi)^{\frac{d}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

2.2 (ii) **4 / 4**

✓ **- 0 pts** Correct

- 1 pts Correct with minor error

- 3 pts Reasonable attempt

- 4 pts Missing / incomplete attempt

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) + (x_2 - b)^T A^{-1}(x_2 - b)\right)$$

iii.

substitute in the given fact:

$$|\Sigma| = |\Sigma_{11}|\left|\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}\right|$$

into our equation for joint density:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1, x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1, x_2)\right)$$

Substitute value of A:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1, x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1, x_2)\right)$$

And split the $Q(x_1, x_2)$ into $Q(x_1) + Q(x_1, x_2)$:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1, x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\,(-\frac{1}{2}(x_1 - \mu_1)\Sigma_{11}^{-1}(x_1 - \mu_1)$$
$$* \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2 - b)^T A^{-1}(x_2 - b)\right)$$

$$f(\vec{x}) = \aleph(x_1, \mu_1, \Sigma_{11}) * \aleph(x_2, b, A)$$

iv.

to find the conditional distribution of x_2 given x_1, divide the joint distribution by the marginal distriubiton for x_1 (from part 1):

$$f_1(x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\,(-\frac{1}{2}(x_1 - \mu_1)\Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\,(-\frac{1}{2}(x_1 - \mu_1)\Sigma_{11}^{-1}(x_1 - \mu_1) * \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2 - b)^T A^{-1}(x_2 - b)\right)$$

2.3 (iii) **4 / 4**

✓ **- 0 pts** Correct

- 1 pts Correct with minor error

- 3 pts Reasonable attempt

- 4 pts Missing / incomplete attempt

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1) + (x_2-b)^T A^{-1}(x_2-b)\right)$$

iii.

substitute in the given fact:

$$|\Sigma| = |\Sigma_{11}|\left|\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right|$$

into our equation for joint density:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right)$$

Substitute value of A:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right)$$

And split the $Q(x_1,x_2)$ into $Q(x_1) + Q(x_1,x_2)$:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}Q(x_1,x_2)\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right)$$
$$* \frac{1}{(2\pi)^{\frac{d}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

$$f(\vec{x}) = \aleph(x_1,\mu_1,\Sigma_{11}) * \aleph(x_2,b,A)$$

iv.

to find the conditional distribution of x_2 given x_1, divide the joint distribution by the marginal distriubiton for x_1 (from part 1):

$$f_1(x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right)$$

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)\right) * \frac{1}{(2\pi)^{\frac{d}{2}}|A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

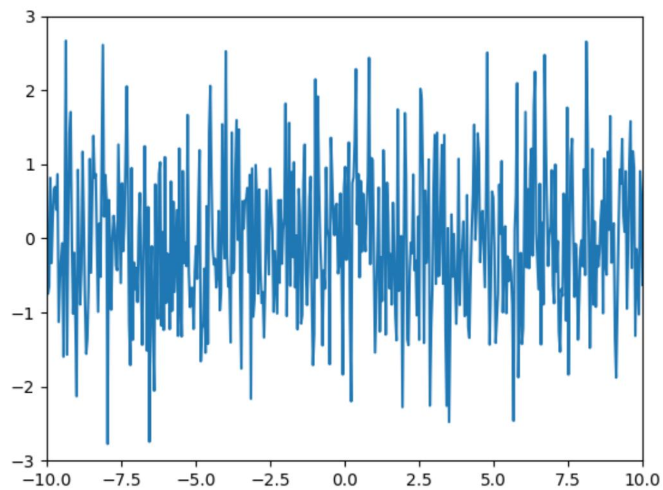
$$f_{2|1}(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)}$$

$$= \frac{\left(\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1) * \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)\right)\right)}{\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}}\exp\,(-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)} =$$

$$f_{2|1}(x_2|x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)\right)$$

v.

Citation: MatplotLib package in python



2.4 (iv) **2 / 2**

- ✓ **+ 2 pts** **Correct answer through marginal**
- **+ 0 pts** Incorrect or Missing
- **+ 1 pts** Almost correct (small typo)

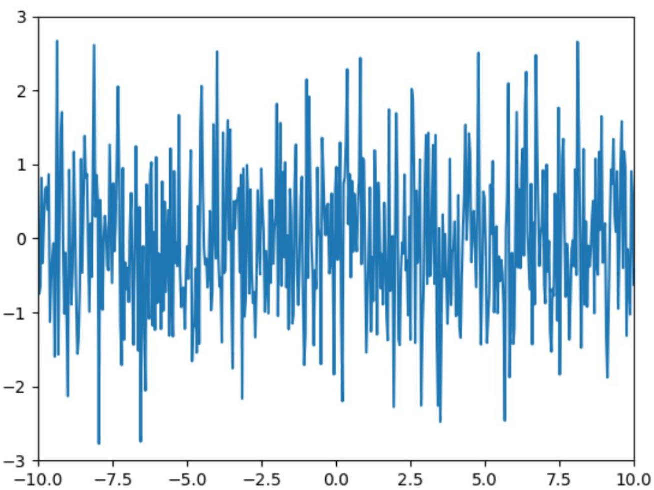
$$f_{2|1}(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)}$$

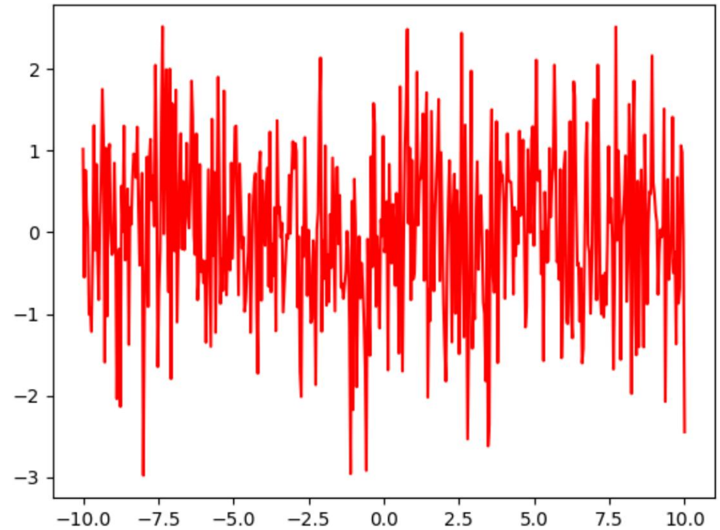
$$= \frac{\left(\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1 - \mu_1)\Sigma_{11}^{-1}(x_1 - \mu_1) * \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2 - b)^T A^{-1}(x_2 - b)\right)\right)\right)}{\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_1 - \mu_1)\Sigma_{11}^{-1}(x_1 - \mu_1)\right)} =$$

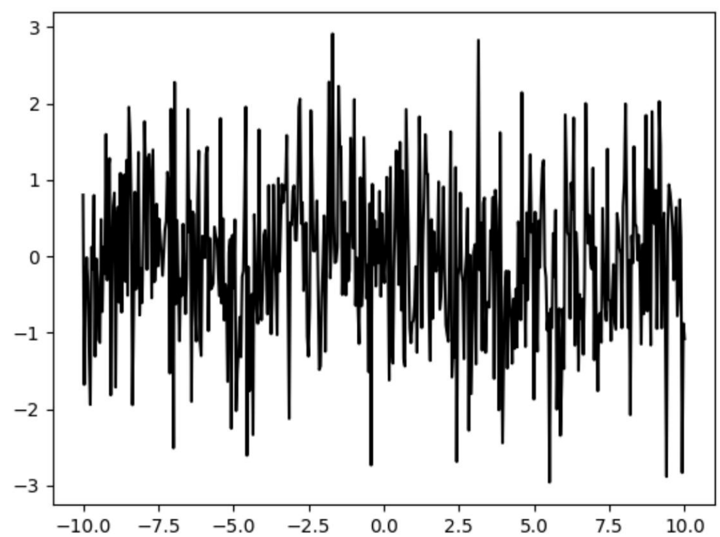
$$f_{2|1}(x_2|x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2 - b)^T A^{-1}(x_2 - b)\right)$$

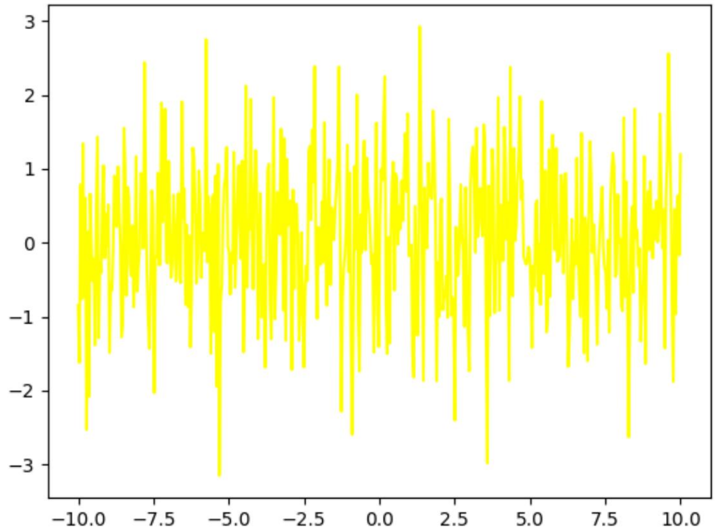
v.

Citation: MatplotLib package in python

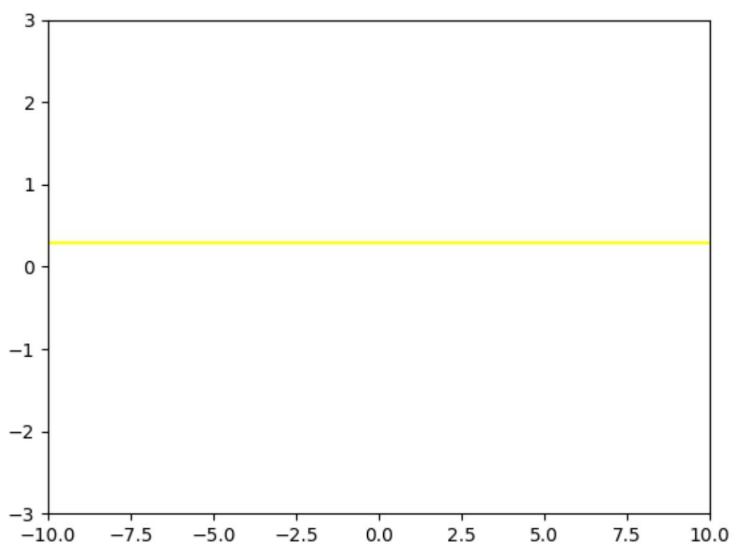



2
1
0
−1
−2
−3
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0


3
2
1
0
−1
−2
−3
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0



Since none of the values are correlated with one another at all (covariance to identity matrix), the function is not smooth. Every x value is random based solely around the mean. When we set covariance to 1, then the plot becomes a straight line (since all the x values are identical to one another because of their relationship w/in the covariance matrix). The y value is constant throughout, but is not neccesarily 0, it is in some sense the "first draw" which determines the y values- eg one of the outputs shown below

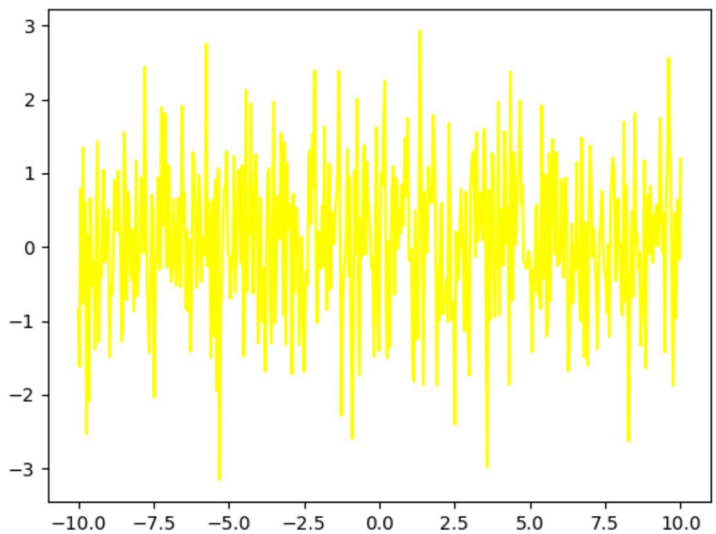


vi.

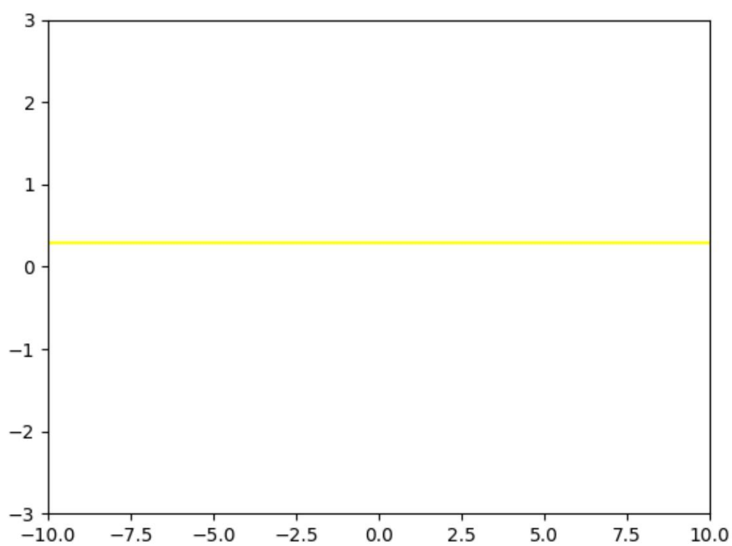
2.5 (v) **4 / 4**

✓ **+ 4 pts** **+1 Plots, +1 smoothness comment, +1 what happens when all ones (constant), +1 some valid explanation about mu AND sigma role OR extra experiment which indicates this trend**

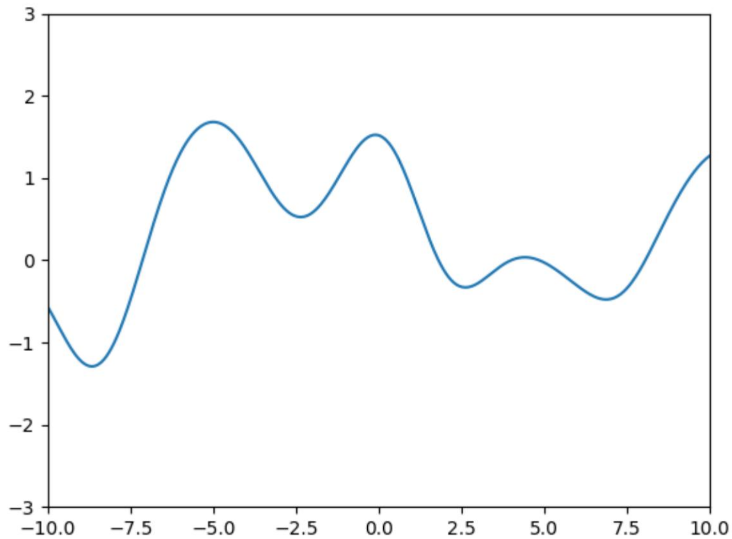
+ **3 pts** 3/4 of above
+ **2 pts** 2/4 of above
+ **1 pts** 1/4 of above
+ **0 pts** Incorrect or missing



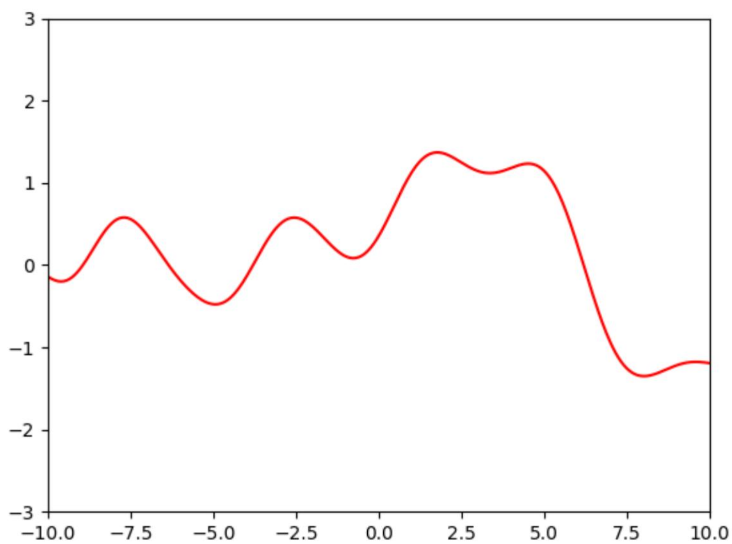
Since none of the values are correlated with one another at all (covariance to identity matrix), the function is not smooth. Every x value is random based solely around the mean. When we set covariance to 1, then the plot becomes a straight line (since all the x values are identical to one another because of their relationship w/in the covariance matrix). The y value is constant throughout, but is not neccesarily 0, it is in some sense the “first draw” which determines the y values- eg one of the outputs shown below



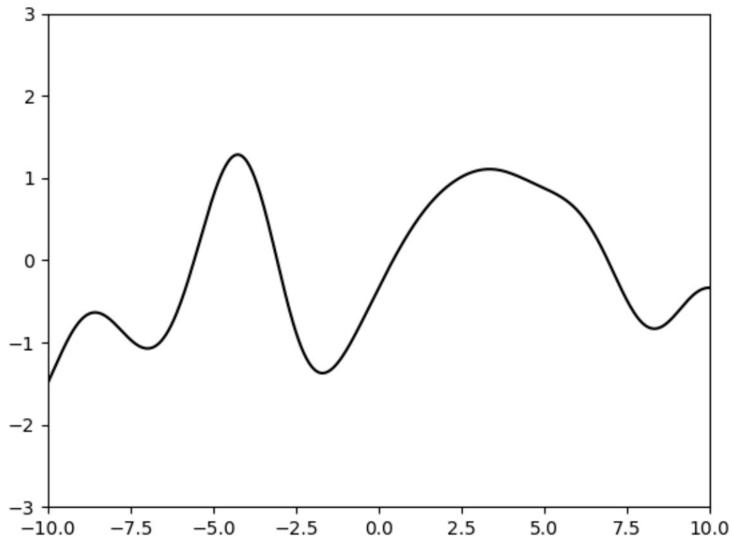
vi.

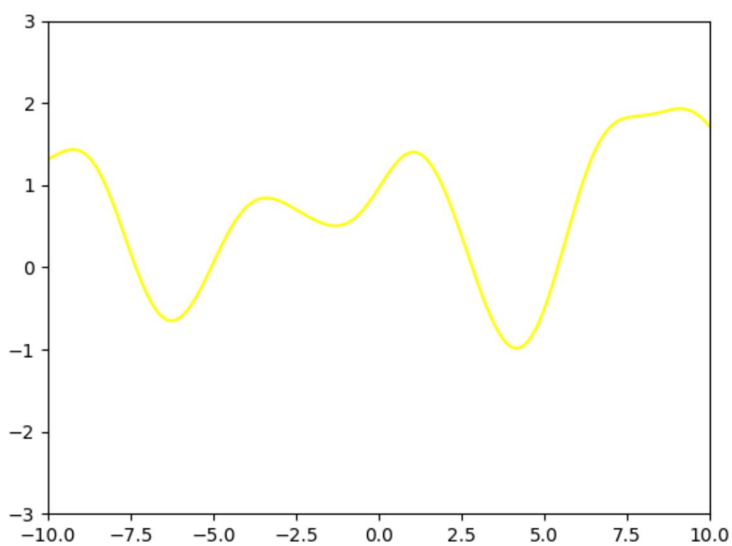


3
2
1
0
−1
−2
−3
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0



3
2
1
0
−1
−2
−3
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0





These functions are guaranteed to be smooth! The points closer to one another are more closely correlated due to the K kernel function which inversely is proportional to the distance between the points.

vii.

In order to generate periodic functions, the covariance must be cyclic – points one full period away must have a covariance of 1. In a general way, a Σ matric which is a function of (x_i, xj) could be any periodic function * distance between x_i, x_j. Choosing to set $\mu - 0,\ \Sigma_{i,j} = \cos\left(\left(x_i - x_j\right)^2 * 2\right)$ produces the below graphs which period of 3 units.

2.6 (vi) **4 / 4**

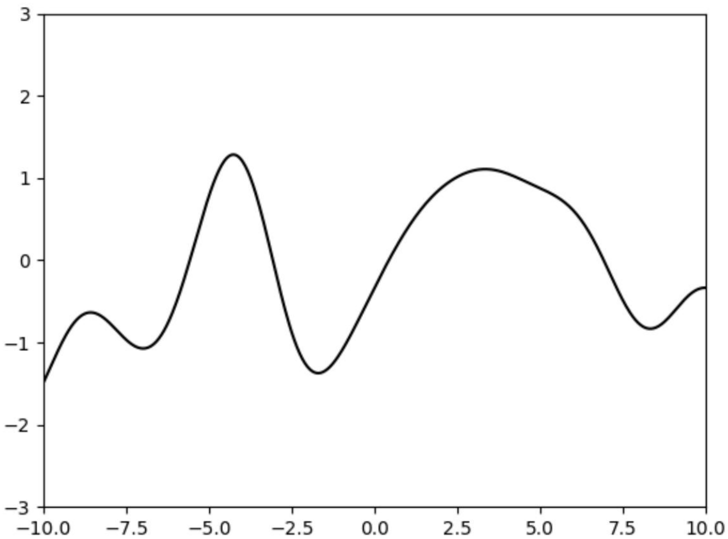
✓ **+ 4 pts** **Correct: BOTH 1 and 2**

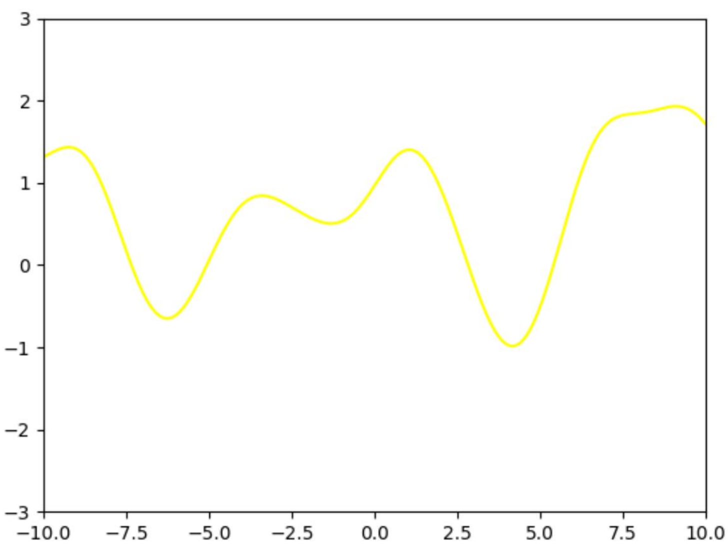
+ 3 pts Either:

1. Only smoothness description AND rationalization but without plots
2. Only plots with description but without reasonable justification for smoothness

+ 1 pts Plots appear right but incorrect/missing description

+ 0 pts Incorrect/Missing

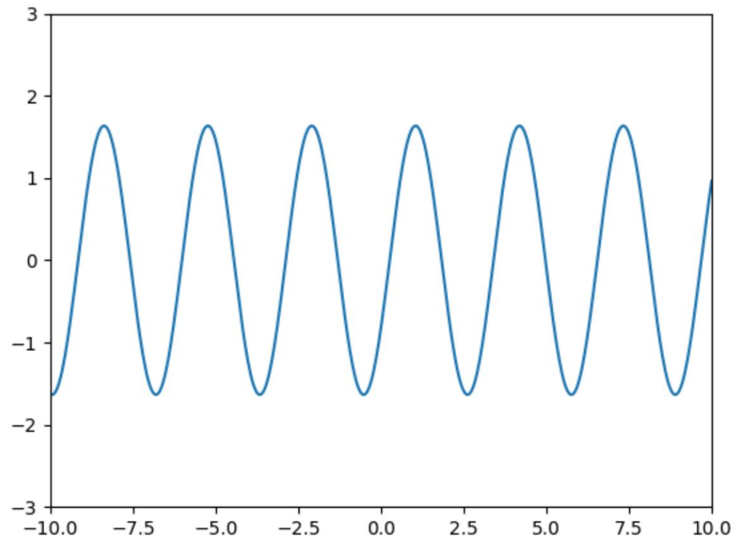


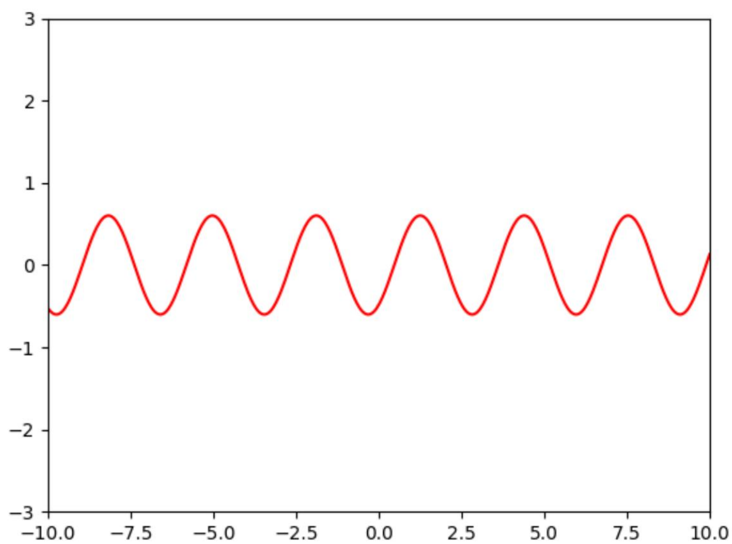


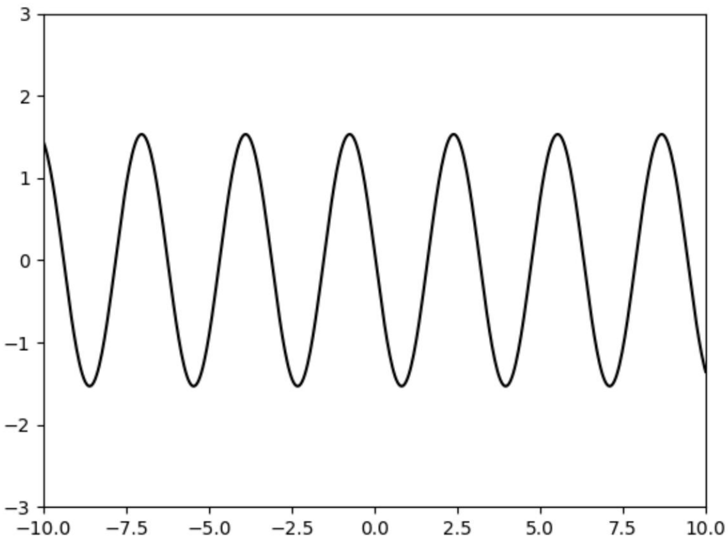
These functions are guaranteed to be smooth! The points closer to one another are more closely correlated due to the K kernel function which inversely is proportional to the distance between the points.

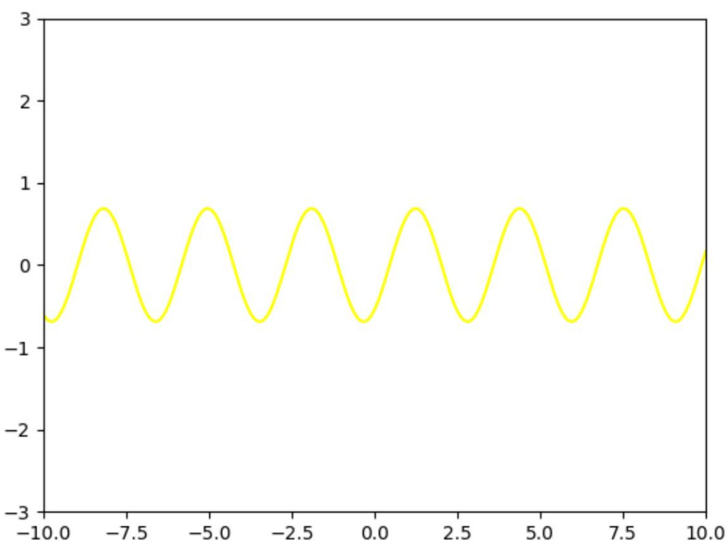
vii.

In order to generate periodic functions, the covariance must be cyclic – points one full period away must have a covariance of 1. In a general way, a Σ matric which is a function of (x_i, xj) could be any periodic function * distance between x_i, x_j. Choosing to set $\mu - 0,\ \Sigma_{i,j} = \cos\left(\left(x_i - x_j\right)^2 * 2\right)$ produces the below graphs which period of 3 units.


3
2
1
0
−1
−2
−3
−10.0 −7.5 −5.0 −2.5 0.0 2.5 5.0 7.5 10.0


3
2
1
0
−1
−2
−3
−10.0 −7.5 −5.0 −2.5 0.0 2.5 5.0 7.5 10.0





viii.

from part iv:

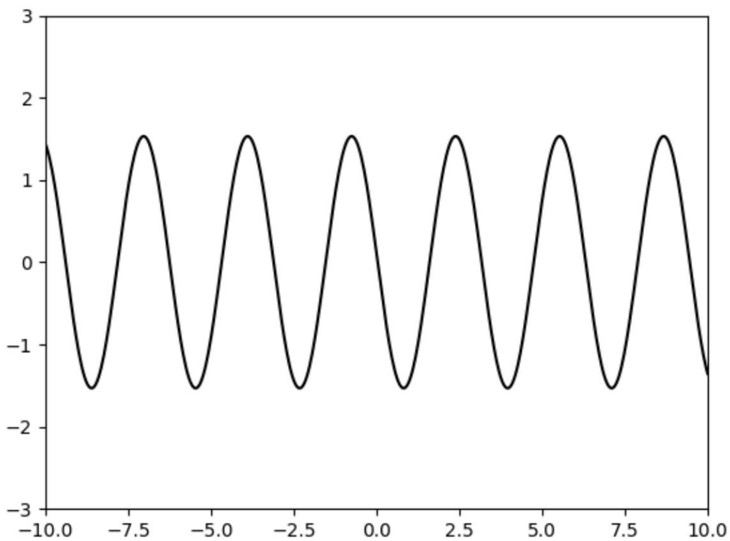
$$f_{2|1}(x_2|x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2 - b)^T A^{-1}(x_2 - b)\right) = \aleph(x_2, b, A)$$

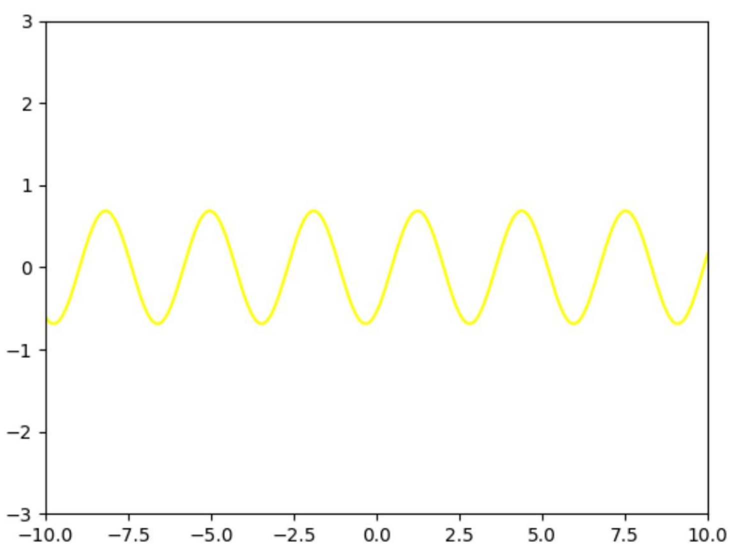
Given the covariance Matrix for Y:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} K(X,X) & K(X,\bar{X}) \\ K(\bar{X},X) & K(\bar{X},\bar{X}) \end{bmatrix}$$

2.7 (vii) **4 / 4**

- ✓ **- 0 pts** Correct
- **- 2 pts** No explanation about settings of $$\Sigma$$ and $$\mu$$
- **- 2 pts** No plot/Incorrect plot
- **- 4 pts** Incorrect/Missing





viii.

from part iv:

$$f_{2|1}(x_2|x_1) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_2 - b)^T A^{-1}(x_2 - b)\right) = \aleph(x_2, b, A)$$

Given the covariance Matrix for Y:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} K(X, X) & K(X, \bar{X}) \\ K(\bar{X}, X) & K(\bar{X}, \bar{X}) \end{bmatrix}$$

However, we need to switch our covariance matrix since what we were given was:

$$\begin{bmatrix} Y \\ \bar{Y} \end{bmatrix} \sim \aleph\left(\begin{bmatrix} \mu_n \\ \mu_m \end{bmatrix}, \begin{bmatrix} K(X,X) & K(X,\bar{X}) \\ K(\bar{X},X) & K(\bar{X},\bar{X}) \end{bmatrix}\right)$$

To:

$$\begin{bmatrix} \bar{Y} \\ \bar{Y} \end{bmatrix} \sim \aleph\left(\begin{bmatrix} \mu_m \\ \mu_n \end{bmatrix}, \begin{bmatrix} K(\bar{X},\bar{X}) & K(\bar{X},X) \\ K(X,\bar{X}) & K(X,X) \end{bmatrix}\right)$$

Then we can apply to our eqn from part iv.

$$f(\Upsilon|\bar{\Upsilon}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\bar{\Upsilon} - b)^T A^{-1}\overline{(\Upsilon} - b)\right)$$

Or:

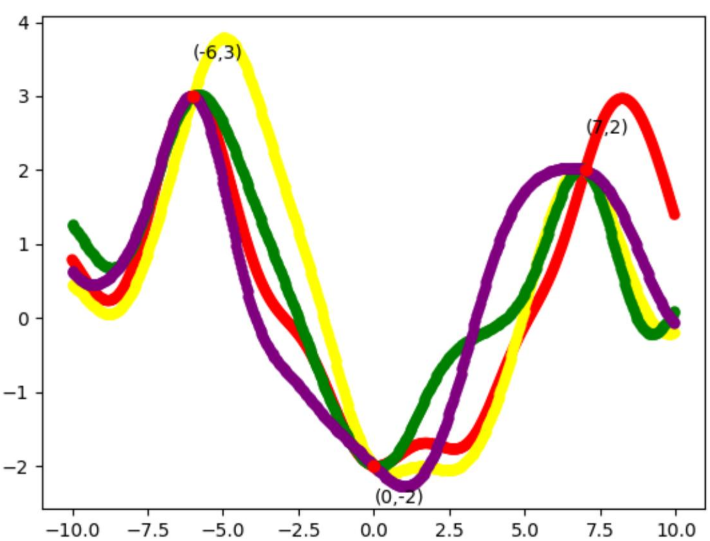
$$f(\Upsilon|\bar{\Upsilon}) = \aleph(x_2, b, A)$$

Where:

$$A = \Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12} = K(X,X) - K^T\overline{(X},X)K^{-1}(\bar{X},\bar{X})K(\bar{X},\bar{X})$$

$$b = \mu_2 - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1) = \mu_n + K^T(\bar{X},X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

ix.



All of the functions pass through the training data points

x.

2.8 (viii) **3 / 3**

- ✓ **- 0 pts** Correct
- **- 1 pts** Minor issue
- **- 3 pts** Incorrect/Missing

However, we need to switch our covariance matrix since what we were given was:

$$\begin{bmatrix} Y \\ \bar{Y} \end{bmatrix} \sim \aleph\left(\begin{bmatrix} \mu_n \\ \mu_m \end{bmatrix}, \begin{bmatrix} K(X,X) & K(X,\bar{X}) \\ K(\bar{X},X) & K(\bar{X},\bar{X}) \end{bmatrix}\right)$$

To:

$$\begin{bmatrix} \bar{Y} \\ \bar{Y} \end{bmatrix} \sim \aleph\left(\begin{bmatrix} \mu_m \\ \mu_n \end{bmatrix}, \begin{bmatrix} K(\bar{X},\bar{X}) & K(\bar{X},X) \\ K(X,\bar{X}) & K(X,X) \end{bmatrix}\right)$$

Then we can apply to our eqn from part iv.

$$f(\Upsilon|\bar{\Upsilon}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\bar{\Upsilon} - b)^T A^{-1}\overline{(\Upsilon} - b)\right)$$

Or:

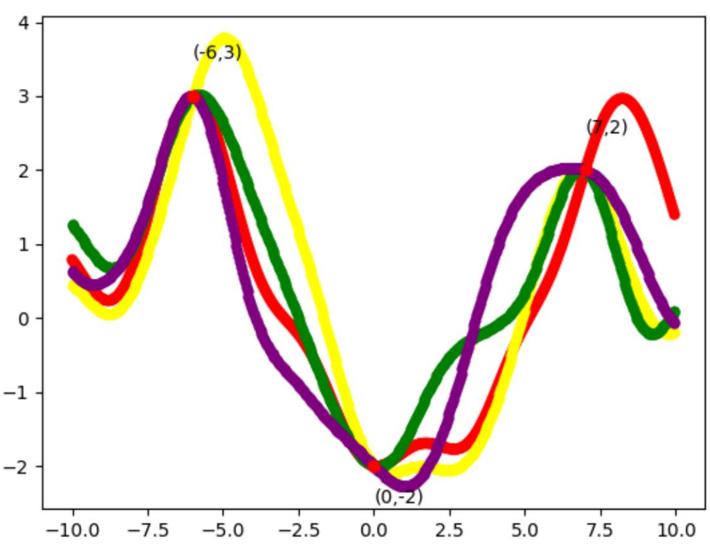
$$f(\Upsilon|\bar{\Upsilon}) = \aleph(x_2, b, A)$$

Where:

$$A = \Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12} = K(X,X) - K^T\overline{(X},X)K^{-1}(\bar{X},\bar{X})K(\bar{X},\bar{X})$$

$$b = \mu_2 - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1) = \mu_n + K^T(\bar{X},X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

ix.



All of the functions pass through the training data points

x.

2.9 (ix) **4 / 4**

✓ **- 0 pts** Correct

- 2 pts Incorrect/Missing written part for this question.

- 2 pts Incorrect/Missing plots(Curves didn't pass through training data or missing training data on plots)

- 4 pts Incorrect/Missing

However, we need to switch our covariance matrix since what we were given was:

$$\begin{bmatrix} Y \\ \bar{Y} \end{bmatrix} \sim \aleph\left(\begin{bmatrix} \mu_n \\ \mu_m \end{bmatrix}, \begin{bmatrix} K(X,X) & K(X,\bar{X}) \\ K(\bar{X},X) & K(\bar{X},\bar{X}) \end{bmatrix}\right)$$

To:

$$\begin{bmatrix} \bar{Y} \\ \bar{Y} \end{bmatrix} \sim \aleph\left(\begin{bmatrix} \mu_m \\ \mu_n \end{bmatrix}, \begin{bmatrix} K(\bar{X},\bar{X}) & K(\bar{X},X) \\ K(X,\bar{X}) & K(X,X) \end{bmatrix}\right)$$

Then we can apply to our eqn from part iv.

$$f(\Upsilon|\bar{\Upsilon}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathrm{A}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\bar{\Upsilon} - b)^T A^{-1}\overline{(\Upsilon} - b)\right)$$

Or:

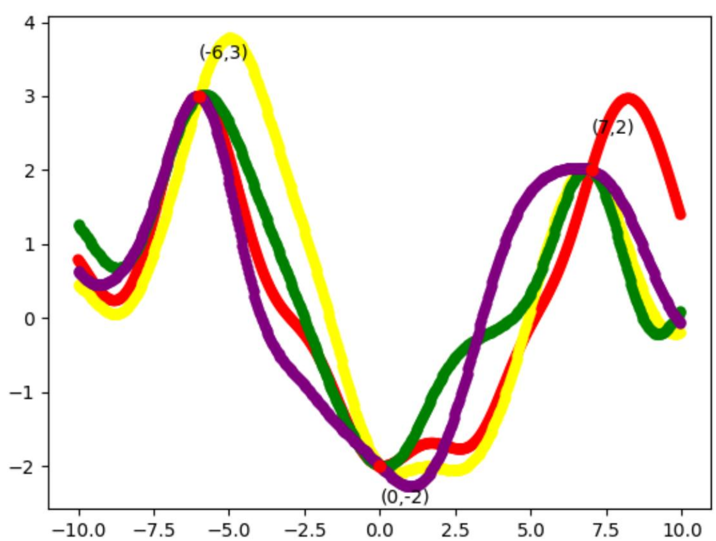
$$f(\Upsilon|\bar{\Upsilon}) = \aleph(x_2, b, A)$$

Where:

$$A = \Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12} = K(X,X) - K^T\overline{(\bar{X},X)}K^{-1}(\bar{X},\bar{X})K(\bar{X},\bar{X})$$

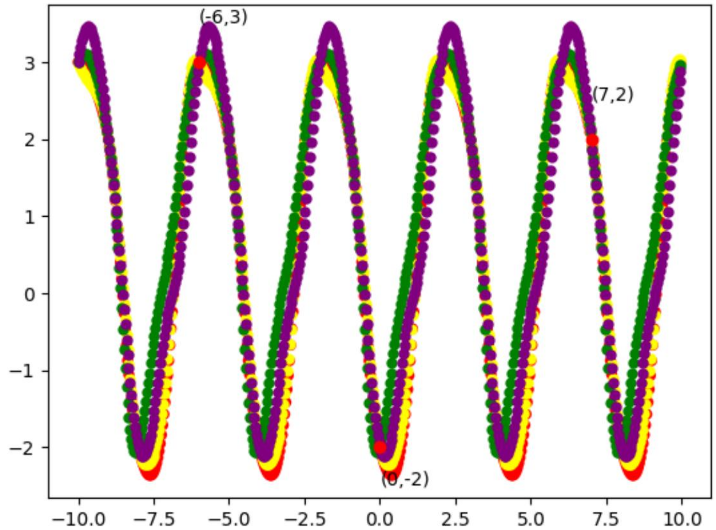
$$b = \mu_2 - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1) = \mu_n + K^T(\bar{X},X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

ix.



All of the functions pass through the training data points

x.



Again, the training data is on the line of each of the four functions. You need to adjust the period also so that it is possible.

xi.

From viii we found the posterior:

$$f(\Upsilon|\overline{\Upsilon}) = \aleph(x_2, b, A)$$

Where:

$$A = \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12} = K(X,X) - K^T\overline{(\bar{X},X)}K^{-1}(\bar{X},\bar{X})K(\bar{X},\bar{X})$$

$$b = \mu_2 - \Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1) = \mu_n + K^T(\bar{X},X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

B is the mean of the posterior:

$$b = \mu_n + K^T(\bar{X},X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

xii.

3. Regression Competition

i., ii. Submitted online

iii.

Resources/Citations:

https://scikit-learn.org/stable/modules/classes.html#module-sklearn.preprocessing

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html

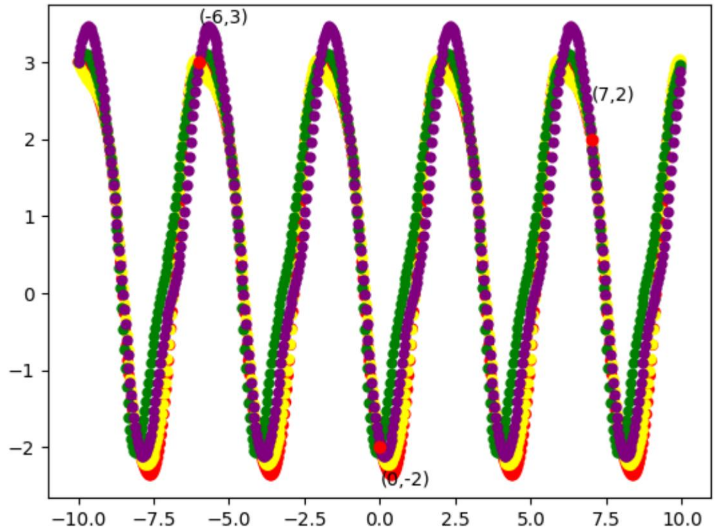
2.10 (x) **4 / 4**

✓ **+ 4 pts** **(+2)4 plots through all training data points with changed period OR plots through some training data using pseudo-inverse (WITH written acknowledgement of issue)**

(+1) overlay of points on graph as proof

(+1) correct written reflection on the graph

- **+ 3 pts** Any three points scored in above rubric
- **+ 2 pts** Any two points scored in above rubric
- **+ 1 pts** Single point scored in above rubric
- **+ 0 pts** Incorrect/missing



Again, the training data is on the line of each of the four functions. You need to adjust the period also so that it is possible.

xi.

From viii we found the posterior:

$$f(\Upsilon|\overline{\Upsilon}) = \aleph(x_2, b, A)$$

Where:

$$A = \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12} = K(X,X) - K^T\overline{(\bar{X},X)}K^{-1}\,(\bar{X},\bar{X})K(\bar{X},\bar{X})$$

$$b = \mu_2 - \Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1) = \mu_n + K^T(\bar{X}\,,X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

B is the mean of the posterior:

$$b = \mu_n + K^T(\bar{X}\,,X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

xii.

3. Regression Competition

i., ii. Submitted online

iii.

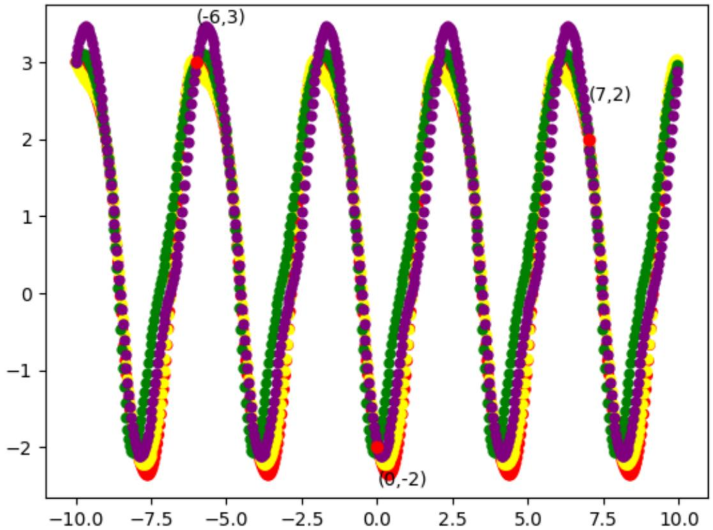
Resources/Citations:

https://scikit-learn.org/stable/modules/classes.html#module-sklearn.preprocessing

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html

2.11 (xi) **2 / 2**

- ✓ **- 0 pts** Correct
- **- 2 pts** Incorrect/Missing



Again, the training data is on the line of each of the four functions. You need to adjust the period also so that it is possible.

xi.

From viii we found the posterior:

$$f(\Upsilon|\overline{\Upsilon}) = \aleph(x_2, b, A)$$

Where:

$$A = \Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12} = K(X,X) - K^T\overline{(\bar{X},X)}K^{-1}(\bar{X},\bar{X})K(\bar{X},\bar{X})$$

$$b = \mu_2 - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1) = \mu_n + K^T(\bar{X},X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

B is the mean of the posterior:

$$b = \mu_n + K^T(\bar{X},X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

xii.

3. Regression Competition

i., ii. Submitted online

iii.

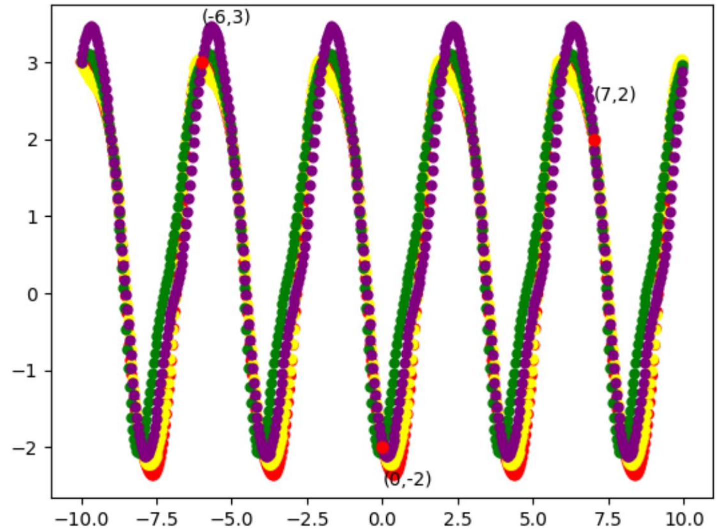
Resources/Citations:

https://scikit-learn.org/stable/modules/classes.html#module-sklearn.preprocessing

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html

2.12 (xii) **0 / 2**

- **- 0 pts** Correct
- **- 1 pts** Incorrect/Missing one plot
- ✓ **- 2 pts Incorrect/Missing**



Again, the training data is on the line of each of the four functions. You need to adjust the period also so that it is possible.

xi.

From viii we found the posterior:

$$f(\Upsilon|\overline{\Upsilon}) = \aleph(x_2, b, A)$$

Where:

$$A = \Sigma_{22} - \Sigma_{12}^{T}\Sigma_{11}^{-1}\Sigma_{12} = K(X,X) - K^{T}\overline{(\bar{X},X)}K^{-1}\,(\bar{X},\bar{X})K(\bar{X},\bar{X})$$

$$b = \mu_2 - \Sigma_{12}^{T}\Sigma_{11}^{-1}(x_1 - \mu_1) = \mu_n + K^{T}(\bar{X}\,,X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

B is the mean of the posterior:

$$b = \mu_n + K^{T}(\bar{X}\,,X)K^{-1}(\bar{X},\bar{X})(\bar{Y} - \mu_m)$$

xii.

3. Regression Competition

i., ii. Submitted online

iii.

Resources/Citations:

https://scikit-learn.org/stable/modules/classes.html#module-sklearn.preprocessing

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html

The regressor with which I obtained the best results and with which I submitted my final predictions was a SDGRegressor from Sci-kit learn using Huber Loss and L1 penalty, and after applying a powerTransform on the dataset. Post-processing, I adjusted using a numpy function any predictions that predicted a year out of the range of the dataset – since we applied a linear regressor, there were predictions of songs from year 2025, e.g. (Although I was very curious what song sounded so futuristic :D). One key choice I made was to move away from using OLS fit, because the evaluation was based on Mean absolute error, not minimizing the squared loss! Similarly, I experimented with using different Scalars – eg the standard Scaler. I found experimentally that the PowerTransformer performed best – one reason for this is because we were given no guarantees that each column was distributed in a guassian way (if this were true, then the standardScaler would have performed better). Other techniques like knn did not perform as well as the linear regressions, and were not helpful. Similarly, I found L1 loss worked better than L2, perhaps because of the nature of our dataset, wherein each song only a few sounds are actually significant.

3 Regression Competition 50 / 50

✓ **- 0 pts** **Comprehensive PDF**

- 5 pts Slightly incomplete/vague explanation [why did you pick a model?/what else did you try?]

- 12.5 pts Incomplete/Vague explanation [what models did you pick, what experiments did you run?]

- 20 pts Minimum PDF submisison credit

- 25 pts No PDF submission found

- 0 pts Click here to replace this description.