DSC 140B - Quiz 05

February 12, 2026

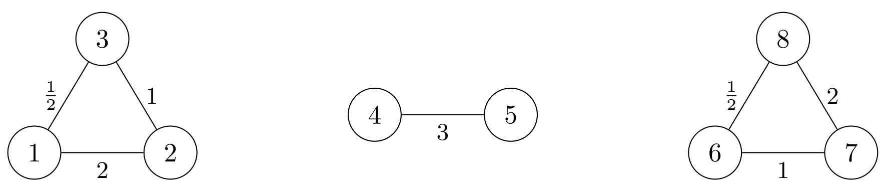
Name: ______ PID: ______

About the quizzes:

- Quizzes in DSC 140B are *optional* and graded pass/fail.
- A score of 70% or higher earns a "pass" and 1.5 credits toward your final grade.
- If you don't pass, no credits are earned, but it doesn't hurt your grade.
- You have 30 minutes to complete the quiz.
- At least one of the questions below will be on an exam (probably with slight changes, such as different numbers).
- Unfortunately, we can't answer clarifying questions during the quiz. If you think a question has a bug or is unclear, please let us known in a private post on Campuswire after the quiz, and we'll take it into account when grading.

Problem 1.

Consider the similarity graph shown below with three connected components:



The edge weights are shown on each edge. Consider the embedding vector:

$$\vec{f} = (3, 3, 3, -1, -1, 5, 5, 5)^T$$

What is the Laplacian Eigenmaps cost of this embedding?

0

Solution: 0.

The Laplacian Eigenmaps cost is given by the double sum:

$$\text{Cost}(\vec{f}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(f_i - f_j)^2$$

Since $w_{ij} = 0$ when there is no edge between nodes i and j, this simplifies to summing only over edges. For each edge (i, j) with weight w_{ij}, the contribution to the cost is $w_{ij}(f_i - f_j)^2$.

Looking at the embedding, nodes in the same connected component have the same embedding value:

- Component 1 (nodes 1, 2, 3): all have $f_i = 3$
- Component 2 (nodes 4, 5): both have $f_i = -1$
- Component 3 (nodes 6, 7, 8): all have $f_i = 5$

Since there are no edges between different components, and within each component all nodes have the same embedding value, every term $(f_i - f_j)^2 = 0$ for connected nodes. Therefore, the total cost is 0.

Problem 2.

Suppose you are given the following basis functions that define a feature map $\vec{\phi} : \mathbb{R}^3 \to \mathbb{R}^4$:

$$\varphi_1(\vec{x}) = x_1^2$$
$$\varphi_2(\vec{x}) = x_2 x_3$$
$$\varphi_3(\vec{x}) = x_1 x_3$$
$$\varphi_4(\vec{x}) = x_1 x_2$$

What is the representation of the data point $\vec{x} = (2, -1, 3)$ in the new feature space?

- ○ $(4, 3, 6, -2)$
- ○ $(4, -3, -6, -2)$
- ● $(4, -3, 6, -2)$
- ○ $(2, -3, 6, -2)$

Solution: $(4, -3, 6, -2)$.

We compute each basis function at $\vec{x} = (2, -1, 3)$:

$$\varphi_1(\vec{x}) = x_1^2 = (2)^2 = 4$$
$$\varphi_2(\vec{x}) = x_2 x_3 = (-1)(3) = -3$$
$$\varphi_3(\vec{x}) = x_1 x_3 = (2)(3) = 6$$
$$\varphi_4(\vec{x}) = x_1 x_2 = (2)(-1) = -2$$

So $\vec{\phi}(\vec{x}) = (4, -3, 6, -2)$.

This was similar to Practice Problem 108.

Problem 3.

Suppose we have a feature map $\varphi : \mathbb{R}^3 \to \mathbb{R}^4$ with the following basis functions:

$$\varphi_1(\vec{x}) = |x_1 + x_2|$$
$$\varphi_2(\vec{x}) = |x_1 - x_3|$$
$$\varphi_3(\vec{x}) = |x_2|$$
$$\varphi_4(\vec{x}) = |x_2 + x_3|$$

A linear prediction function in this feature space has learned the weight vector $\vec{w} = (w_0, w_1, w_2, w_3, w_4) = (1,\ 2,\ -3,\ 1,\ -1)$, where $w_0 = 1$ is the bias (intercept) term. The prediction function is:

$$H(\vec{x}) = w_0 + w_1\varphi_1(\vec{x}) + w_2\varphi_2(\vec{x}) + w_3\varphi_3(\vec{x}) + w_4\varphi_4(\vec{x})$$

What is the value of $H(\vec{x})$ for the input point $\vec{x} = (2, -4, 1)$?

3

Solution: 3.

First, we compute the feature representation of $\vec{x} = (2, -4, 1)$:

$$\begin{aligned}
\varphi_1(\vec{x}) &= |x_1 + x_2| = |2 + (-4)| = |-2| = 2 \\
\varphi_2(\vec{x}) &= |x_1 - x_3| = |2 - 1| = 1 \\
\varphi_3(\vec{x}) &= |x_2| = |-4| = 4 \\
\varphi_4(\vec{x}) &= |x_2 + x_3| = |-4 + 1| = |-3| = 3
\end{aligned}$$

So the feature vector is $\varphi(\vec{x}) = (2, 1, 4, 3)$.

Then we compute the prediction function:

$$\begin{aligned}
H(\vec{x}) &= w_0 + w_1\varphi_1(\vec{x}) + w_2\varphi_2(\vec{x}) + w_3\varphi_3(\vec{x}) + w_4\varphi_4(\vec{x}) \\
&= 1 + 2(2) + (-3)(1) + 1(4) + (-1)(3) \\
&= 1 + 4 - 3 + 4 - 3 \\
&= 3
\end{aligned}$$

You could also compute this directly by substituting the definitions of the basis functions into the prediction function:

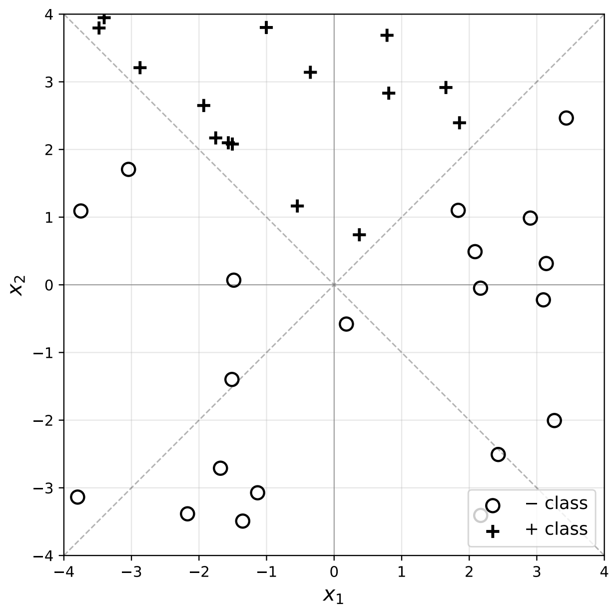
$$\begin{aligned}
H(\vec{x}) &= w_0 + w_1\varphi_1(\vec{x}) + w_2\varphi_2(\vec{x}) + w_3\varphi_3(\vec{x}) + w_4\varphi_4(\vec{x}) \\
&= 1 + 2|x_1 + x_2| - 3|x_1 - x_3| + 1|x_2| - 1|x_2 + x_3| \\
&= 1 + 2|2 + (-4)| - 3|2 - 1| + 1|-4| - 1|-4 + 1| \\
&= 1 + 2(2) - 3(1) + 1(4) - 1(3) \\
&= 3
\end{aligned}$$

Either way works!

This was similar to Practice Problem 113.

Problem 4.

Consider the data shown below:



The data comes from two classes: $\circ$ (negative) and $+$ (positive). The dashed lines represent $y = x$ and $y = -x$.

Suppose a single basis function will be used to map the data to a 1-dimensional feature space where a linear classifier will be trained. Which of the below is the best choice of basis function?

- ○ $\varphi(x_1, x_2) = x_1 + x_2$
- ○ $\varphi(x_1, x_2) = x_1 \cdot x_2$
- ● $\varphi(x_1, x_2) = x_2 - |x_1|$
- ○ $\varphi(x_1, x_2) = x_1^2 + x_2^2$

Solution: $\varphi(x_1, x_2) = x_2 - |x_1|$.

There are a couple of ways to reason out which choice of basis function is best.

One way is to try and understand the geometry of the data. Here, the positive class points lie in the region where $x_2 > |x_1|$ (above both the lines $x_2 = x_1$ and $x_2 = -x_1$). The negative class points lie where $x_2 < |x_1|$.

The basis function $\varphi(x_1, x_2) = x_2 - |x_1|$ maps:

- Positive class points to positive values (since $x_2 > |x_1|$ implies $x_2 - |x_1| > 0$)
- Negative class points to negative values (since $x_2 < |x_1|$ implies $x_2 - |x_1| < 0$)

This allows a simple threshold at 0 in the 1D feature space to perfectly separate the classes.

Another way is to test each basis function to see whether it is capable of separating the classes in the 1D feature space. Since we only had four choices in this problem, we can essentially "guess and check" each one by taking one or more "test points" from each class and ensuring that the basis function separates them in the 1D feature space.

- The first option, $\varphi(x_1, x_2) = x_1 + x_2$, won't work well: a point at $(0, 2)$ and a point at $(2, 0)$ should belong to the positive and negative classes, respectively, but they would both be mapped to the same value of 2.

- The second option, $\varphi(x_1, x_2) = x_1 \cdot x_2$, also won't work well: a point at $(0, 2)$ and a point at $(2, 0)$ would both be mapped to 0.
- The third option, $\varphi(x_1, x_2) = x_2 - |x_1|$, works well for our test points at $(0, 2)$ and a point at $(2, 0)$, since they would be mapped to 2 and -2, respectively. This doesn't mean it is a good basis function, since it might not work for other test points, but it's still a possibility.
- The fourth option, $\varphi(x_1, x_2) = x_1^2 + x_2^2$, won't work well: a point at $(0, 2)$ and a point at $(2, 0)$ would both be mapped to 4.

Since the third option is the only one that works well for our test points, it is the best choice of basis function. And in fact, it works well for all points in the dataset for the reasons described in the first part of the solution.

This was similar to Practice Problem 115.

Problem 5. (3 points)

Define the "box" basis function:

$$\phi(x; c) = \begin{cases} 1, & |x - c| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Three box basis functions ϕ_1, ϕ_2, ϕ_3 have centers $c_1 = 0$, $c_2 = 3$, and $c_3 = 5$, respectively. These basis functions map data from $\mathbb{R}$ to feature space $\mathbb{R}^3$ via $x \mapsto (\phi_1(x), \phi_2(x), \phi_3(x))^T$.

A linear predictor in feature space has equation:

$$H_\phi(\vec{z}) = 4z_1 + 2z_2 - 3z_3$$

a) What is the representation of $x = 4$ in feature space?

- ○ $(1, 1, 0)^T$
- ● $(0, 1, 1)^T$
- ○ $(0, 1, 0)^T$
- ○ $(1, 0, 1)^T$

Solution: $(0, 1, 1)^T$.

We evaluate each basis function at $x = 4$:

$$\begin{aligned} \phi_1(4) &= 0, \quad \text{since } |4 - 0| = 4 > 1 \\ \phi_2(4) &= 1, \quad \text{since } |4 - 3| = 1 \leq 1 \\ \phi_3(4) &= 1, \quad \text{since } |4 - 5| = 1 \leq 1 \end{aligned}$$

Therefore, the feature space representation is $(0, 1, 1)^T$.

This was similar to Practice Problem 118.

b) What is $H(2)$?

2

Solution: 2.

First, we find the feature space representation of $x = 2$:

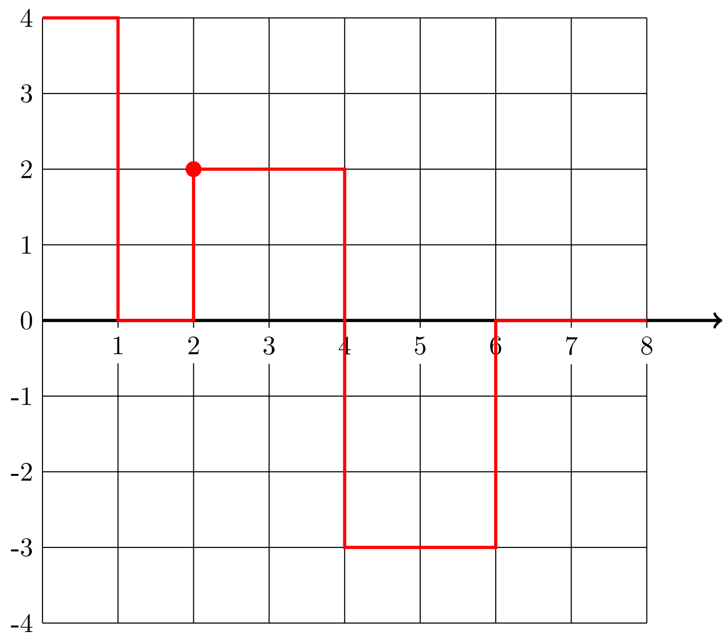
$$\begin{aligned}\phi_1(2) &= 0, \quad \text{since } |2-0| = 2 > 1\\ \phi_2(2) &= 1, \quad \text{since } |2-3| = 1 \le 1\\ \phi_3(2) &= 0, \quad \text{since } |2-5| = 3 > 1\end{aligned}$$

Then:

$$\begin{aligned}H(2) &= H_\phi(0,1,0)\\ &= 4(0) + 2(1) - 3(0)\\ &= 2\end{aligned}$$

This was similar to Practice Problem 118.

c) Plot $H(x)$ (the prediction function in the original space) from 0 to 8 on the grid below.



Solution: We can think of the ith basis function as a box, centered at c, with width 1 and height w_i. The prediction function $H(x)$ is the sum of the heights of the boxes that contain x.

The first basis function box has a height of 4, the second has a height of 2, and the third has a height of -3. The first box covers the interval $[-1, 1]$, the second box covers the interval $[2, 4]$, and the third box covers the interval $[4, 6]$. This results in the plot shown above.

The point at $x = 2$ is highlighted in red since it was the answer to the previous part of the problem. The previous part essentially asked you to evaluate H at a single point, and this part asked you to evaluate H at many points and plot the result. Your two answers should hopefully be consistent with each other!

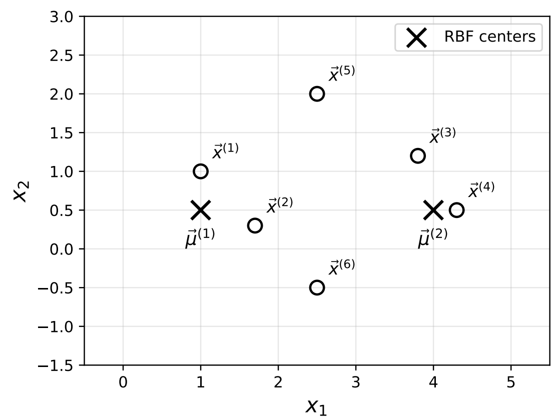
Finally, in practice we usually don't use box basis functions (we use Gaussian basis functions instead), but the same basic principles apply when understanding how Gaussian RBF networks work.

Note: if we were being very careful, we would draw open and closed circles at the jump discontinuities to precisely indicate the value of H at those points, but we won't worry about being so precise here.

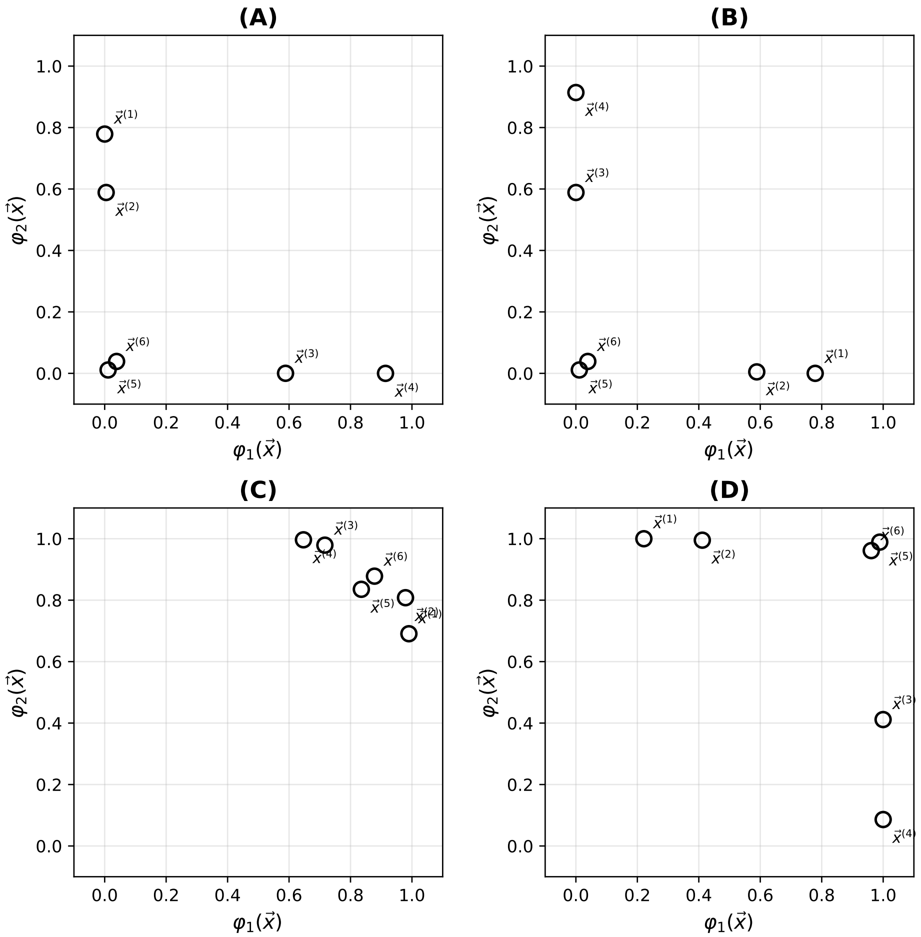
This was Practice Problem 118 with different weights and RBF locations.

Problem 6.

Consider 6 data points $\vec{x}^{(1)}, \ldots, \vec{x}^{(6)}$ shown below in the original feature space, along with two Gaussian RBF basis function centers $\vec{\mu}^{(1)}$ and $\vec{\mu}^{(2)}$ (shown as x markers). Both RBF basis functions use $\sigma = 1$.



One of the plots below shows the data after it has been mapped to feature space using the two Gaussian RBF basis functions centered at $\vec{\mu}^{(1)}$ and $\vec{\mu}^{(2)}$. Which plot is it?



○ (A) ● (B) ○ (C) ○ (D)

Solution: (B).

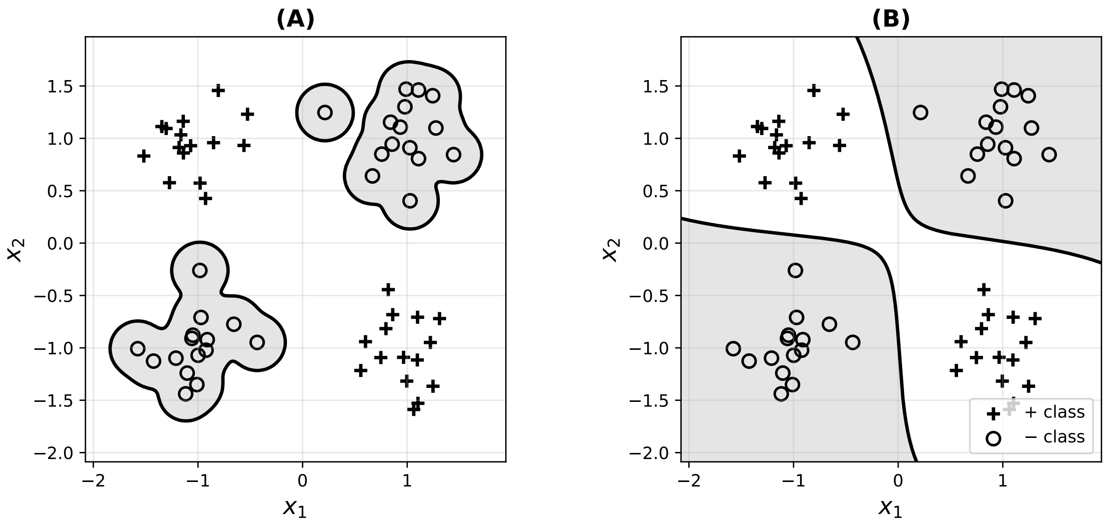
Points close to $\vec{\mu}^{(1)}$ should have high φ_1 values (close to 1) and low φ_2 values (close to 0). Points close to $\vec{\mu}^{(2)}$ should have the opposite. Points far from both centers should have both values close to 0.

- $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ are close to $\vec{\mu}^{(1)}$ but far from $\vec{\mu}^{(2)}$, so they should appear in the bottom-right region (high φ_1, low φ_2).
- $\vec{x}^{(3)}$ and $\vec{x}^{(4)}$ are close to $\vec{\mu}^{(2)}$, but far from $\vec{\mu}^{(1)}$, so they should appear in the top-left region (low φ_1, high φ_2).
- $\vec{x}^{(5)}$ and $\vec{x}^{(6)}$ are roughly equidistant from both centers, so they should have similar (and small) values for both φ_1 and φ_2.

Only plot (B) shows this pattern correctly. Plot (A) has the axes swapped. Plot (C) shows a transformation that plausibly could result if σ were much larger. Plot (D) shows an "inversion" of the pattern: the *further* a point is from a center, the higher its value for the corresponding basis function, which is not how Gaussian RBFs work.

Problem 7.

Below are two decision boundaries produced by Gaussian RBF classifiers trained on the same data. Both classifiers use the same RBF basis function centers, but with different width parameters σ.



Which classifier uses a larger value of σ?

○ (A)

● (B)

Solution: (B).

The Gaussian RBF basis function is $\varphi(\vec{x}; \vec{\mu}) = \exp\left(-\frac{\|\vec{x}-\vec{\mu}\|^2}{\sigma^2}\right)$.

In lecture, we described a Gaussian basis function as a "bump" centered at $\vec{\mu}$, where the width of the bump is determined by σ. A larger value of σ corresponds to a wider bump, whereas a smaller value of σ corresponds to a narrower bump (more like a spike). The prediction function's surface is found by summing up the bumps corresponding to each basis function, weighted by the learned weights. The decision boundary is where the prediction function's surface has a height of zero.

In the plots above, the first one seems to have come from spikier bumps (i.e., smaller σ). In fact, we can kind of see where individual bumps are located in the first plot, but not in the second.

In general, bigger σ will lead to smoother decision boundaries, and smaller σ will lead to more complex decision boundaries that can fit the training data more closely (but might not generalize as well to test data). In the language of DSC 80 and DSC 140A: smaller sigma leads to greater model complexity and therefore a greater risk of overfitting.

Problem 8.

Does standardizing features affect the output of a Gaussian RBF network?

More precisely, let $\mathcal{X} = \{(\vec{x}^{(1)}, y_1), \ldots, (\vec{x}^{(n)}, y_n)\}$ be a dataset, and let $\mathcal{Z} = \{(\vec{z}^{(1)}, y_1), \ldots, (\vec{z}^{(n)}, y_n)\}$ be the dataset obtained by standardizing each feature in $\mathcal{X}$ (that is, each feature is scaled to have unit standard deviation and zero mean). Note that the y_i's are not standardized – they are the same in both datasets.

Suppose $H_{\mathcal{X}}(\vec{x})$ is a Gaussian RBF network trained by minimizing mean squared error on $\mathcal{X}$, and $H_{\mathcal{Z}}(\vec{z})$ is a Gaussian RBF network trained by minimizing mean squared error on $\mathcal{Z}$. Both RBF networks use the same width parameters, σ. The centers used in $H_{\mathcal{Z}}$ are obtained by standardizing the centers used in $H_{\mathcal{X}}$

using the mean and standard deviation of the data, $\vec{x}^{(i)}$.

True or False: it must be the case that $H_{\mathcal{X}}(\vec{x}^{(i)}) = H_{\mathcal{Z}}(\vec{z}^{(i)})$ for all $i \in \{1, \ldots, n\}$.

◯ True

● False

Solution: False.

Standardizing changes the distances between points and centers. The Gaussian basis function $\varphi(\vec{x}; \vec{\mu}, \sigma) = e^{-\|\vec{x}-\vec{\mu}\|^2/\sigma^2}$ depends on the Euclidean distance $\|\vec{x} - \vec{\mu}\|$. While the centers are standardized along with the data, the width parameter σ remains the same. Since standardizing changes the scale of the features (and thus the distances), using the same σ will produce different basis function outputs, leading to different predictions.

This was Practice Problem 121, verbatim.