# Unit 15 - Week 12 - Assorted topics



## Assignment 12

**The due date for submitting this assignment has passed.**
**As per our records you have not submitted this assignment.**

**Due on 2019-10-23, 23:59 IST.**

Instructions:

1. **Attempt all questions.**
2. **Submission deadline**: 23rd October 2019 23:59 IST
3. **Solutions to be posted**: 24th October 2019
4. Older browsers might show **unnecessary vertical bars** at the end of math equations.

1) Let $U_k$ denote the $d \times k$ matrix of the top $k$ eigenvectors of the covariance matrix ($U_k$ is a truncated version of $U$, which is the matrix of eigenvectors of the covariance matrix). ***1.5 points***

There are two approaches to computing the low-dimensional reconstruction $w \in \mathbb{R}^k$ of a data point $x \in \mathbb{R}^d$ :

- Solve a least squares problem to minimize the reconstruction error.
- Project x onto the span of the columns of $U_k$.

We would like to find $w$ such that the reconstruction error is minimum. The objective function is

- $\min_{w} \|U_k w + x_i\|^2$
- $\min_{U_k} \|U_k w - x_i\|^2$
- $\min_{U_k} \|U_k w + x_i\|^2$
- $\min_{w} \|U_k w - x_i\|^2$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$\min_{w} \|U_k w - x_i\|^2$

2) In continuation with question 1, the solution $w^\star$ is given by ***1 point***

- $(U_k U_k^{\mathrm{T}})^{-1} U_k^{\mathrm{T}} x_i$
- $(U_k^{\mathrm{T}} U_k)^{-1} U_k^{\mathrm{T}} x_i$
- $(U_k^{\mathrm{T}} U_k)\, U_k^{\mathrm{T}} x_i$
- $(U_k^{\mathrm{T}} U_k)^{-1} U_k x_i$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$(U_k^{\mathrm{T}} U_k)^{-1} U_k^{\mathrm{T}} x_i$

3) Consider a $d$-dimensional real valued data $\{x_i\}_{i=1}^{N}$ and feature mapping $\phi : \mathbb{R}^d \to \mathbb{R}^m$ and assume that the data is centered in the feature space i.e., $\sum_{i=1}^{N} \phi(x_i) = 0$ In the following $\Phi$ is a design matrix whose $i^{\text{th}}$ row is $\phi(x_i)$. Recall that as part of PCA, we must solve the eigenvalue problem $Sv = \lambda v$ where $S$ proportional to the sample covariance matrix. For PCA in feature space, we have $S = \sum_{i=1}^{N} \phi(x_i)\phi(x_i)^{\mathrm{T}}$. Why is this a problem if m is large? ***2 points***

- Working in feature space directly is too expensive if $m$ is large.
- The covariance matrix $\Phi^{\mathrm{T}}\Phi$ is $m \times m$, which is too large to compute and work with.
- For large $m$, the inverse of $\Phi$ does not exist always.
- For large $m$, the largest eigenvalue of $S$ tends to $\infty$.

No, the answer is incorrect.
Score: 0

Accepted Answers:
*Working in feature space directly is too expensive if $m$ is large.*
*The covariance matrix $\Phi^{\mathrm{T}}\Phi$ is $m \times m$, which is too large to compute and work with.*

4) In continuation with question 3, you are given a kernel function $k(x, x') = \phi(x).\phi(x')$. Define the kernel matrix $K_{ij} = k(x_i, x_j)$. Choose the correct answer. ***2 points***

- If $\lambda \neq 0$, then $\lambda$ is an eigenvalue of $S$ if and only if $\lambda$ is also an eigenvalue of $K$.
- If $\lambda = 0$, then $\lambda$ is an eigenvalue of $S$ if and only if $\lambda$ is also an eigenvalue of $K$.
- If $\lambda \neq 0$, then $\lambda$ is an eigenvalue of $S$ if $\lambda$ is also an eigenvalue of $K$.
- If $\lambda \neq 0$, then $\lambda$ is an eigenvalue of $S$ if $\lambda$ is not an eigenvalue of $K$

No, the answer is incorrect.
Score: 0

Accepted Answers:
*If $\lambda \neq 0$, then $\lambda$ is an eigenvalue of $S$ if and only if $\lambda$ is also an eigenvalue of $K$.*

5) In continuation with question 3 and 4, let $v$ be an eigenvector of $S$ with nonzero eigenvalue $\lambda$. Let $\alpha_v = \frac{\alpha}{\lambda}$. Then $v$ can be written as ***1 point***

- $v = \frac{1}{\|\phi(x_i)\|}\Phi^{\mathrm{T}}\alpha_v$ where $\alpha_v$ is an eigenvector of $K$ with eigenvalue $\lambda$.
- $v = \Phi^{\mathrm{T}}\alpha_v$ where $\alpha_v$ is an eigenvector of $K$ with eigenvalue $\lambda$.
- $v = \frac{1}{\|\phi(x_i)\|}\Phi^{-1}\alpha_v$ where $\alpha_v$ is an eigenvector of $K$ with eigenvalue $\lambda$.
- $v = \Phi^{-1}\alpha_v$ where $\alpha_v$ is an eignevector of $K$ with eigenvalue $\lambda$.

No, the answer is incorrect.
Score: 0

Accepted Answers:
*$v = \Phi^{\mathrm{T}}\alpha_v$ where $\alpha_v$ is an eigenvector of $K$ with eigenvalue $\lambda$.*

6) In continuation with question 3, 4 and 5, you are given a new data point $x \in \mathbb{R}^d$. The scalar projection of its feature representation $\phi(x)$ onto $\frac{v}{\|v\|}$ (with $v$ defined as above) is ***1.5 points***

- $\frac{\alpha_v^{\mathrm{T}}\Phi}{\|\alpha_v\|}$
- $\frac{\alpha_v^{\mathrm{T}}\Phi\phi(x)}{\sqrt{\lambda}\|\alpha_v\|}$
- $\frac{\alpha_v^{\mathrm{T}}\Phi\phi(x)}{\lambda\|\alpha_v\|}$
- $\frac{\alpha_v^{\mathrm{T}}\Phi\phi(x)}{\|\alpha_v\|}$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$\frac{\alpha_v^{\mathrm{T}}\Phi\phi(x)}{\sqrt{\lambda}\|\alpha_v\|}$

7) A single neuron with a normalized Hebbian-type adaptation rule ***1 point***

- Can extract the first principal component of the data distinction.
- Is a stable non-linear dynamical system.
- Can extract all the principal components of the data distinction.
- Is not a stable non-linear dynamical system.

No, the answer is incorrect.
Score: 0

Accepted Answers:
*Can extract the first principal component of the data distinction.*
*Is a stable non-linear dynamical system.*

8) Recall the non-linear stochastic differential equation we did in the lectures ***2 points***

$$\underline{w}(n+1) = \underline{w}(n) + \eta\left(\underline{x}(n)\underline{x}(n)^{\mathrm{T}}\underline{w}(n) - \underline{w}(n)^{\mathrm{T}}\underline{x}(n)\underline{x}(n)^{\mathrm{T}}\underline{w}(n)\underline{w}(n)\right)$$

Define the "characteristic" matrix of the algorithm as that matrix which when multiplied by the old weight $\underline{w}(n)$ yields the new updated weight vector $\underline{w}(n+1)$. The characteristic matrix for the system described in the above equation is

- $M = I + \eta\left(\underline{x}(n)\underline{x}(n)^{\mathrm{T}} - \underline{w}(n)^{\mathrm{T}}\underline{x}(n)\underline{x}(n)^{\mathrm{T}}\underline{w}(n)\right)$
- $M = I - \eta\left(\underline{x}(n)\underline{x}(n)^{\mathrm{T}} - \underline{w}(n)^{\mathrm{T}}\underline{x}(n)\underline{x}(n)^{\mathrm{T}}\underline{w}(n)\right)$
- $M = I + \eta\left(\underline{x}(n)\underline{x}(n)^{\mathrm{T}} + \underline{w}(n)^{\mathrm{T}}\underline{x}(n)\underline{x}(n)^{\mathrm{T}}\underline{w}(n)\right)$
- $M = I - \eta\left(\underline{x}(n)\underline{x}(n)^{\mathrm{T}} + \underline{w}(n)^{\mathrm{T}}\underline{x}(n)\underline{x}(n)^{\mathrm{T}}\underline{w}(n)\right)$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$M = I + \eta\left(\underline{x}(n)\underline{x}(n)^{\mathrm{T}} - \underline{w}(n)^{\mathrm{T}}\underline{x}(n)\underline{x}(n)^{\mathrm{T}}\underline{w}(n)\right)$

9) Define $R = \mathbb{E}\left(\underline{x}(n)\underline{x}(n)^{\mathrm{T}}\right)$. Compute $\mathbb{E}(M)$, where the expectation is over the input distribution. ***1 point***

- $\mathbb{E}(M) = I + \left(R - \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)$
- $\mathbb{E}(M) = I - \left(R - \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)$
- $\mathbb{E}(M) = I + \left(R + \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)$
- $\mathbb{E}(M) = I - \left(R + \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$\mathbb{E}(M) = I + \left(R - \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)$

10) What is the simplified form of the deterministic non-linear stochastic differential equation obtained by averaging equation given in question 8 over the data point density? ***1 point***

- $\underline{w}(n+1) = \underline{w}(n) + \eta\left(R - \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)\underline{w}(n)$
- $\underline{w}(n+1) = \underline{w}(n) - \eta\left(R - \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)\underline{w}(n)$
- $\underline{w}(n+1) = \underline{w}(n) + \eta\left(R + \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)\underline{w}(n)$
- $\underline{w}(n+1) = \underline{w}(n) - \eta\left(R + \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)\underline{w}(n)$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$\underline{w}(n+1) = \underline{w}(n) + \eta\left(R - \underline{w}(n)^{\mathrm{T}}R\underline{w}(n)I\right)\underline{w}(n)$

11) Define $\Delta\underline{w}(n) = \underline{w}(n+1) - \underline{w}(n)$. What is the governing ordinary non-linear differential equation? ***2 points***

- $\frac{d\underline{w}(t)}{dt} = R\underline{w}(t) - \left(\underline{w}(t)^{\mathrm{T}}R\underline{w}(t)\right)\underline{w}(t)$
- $\frac{d\underline{w}(t)}{dt} = R\underline{w}(t) + \left(\underline{w}(t)^{\mathrm{T}}R\underline{w}(t)\right)\underline{w}(t)$
- $\frac{d\underline{w}(t)}{dt} = R\underline{w}(t) + \left(\underline{w}(t)R^{-1}\underline{w}(t)\right)\underline{w}(t)$
- $\frac{d\underline{w}(t)}{dt} = R\underline{w}(t) - \left(\underline{w}(t)R^{-1}\underline{w}(t)\right)\underline{w}(t)$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$\frac{d\underline{w}(t)}{dt} = R\underline{w}(t) - \left(\underline{w}(t)^{\mathrm{T}}R\underline{w}(t)\right)\underline{w}(t)$

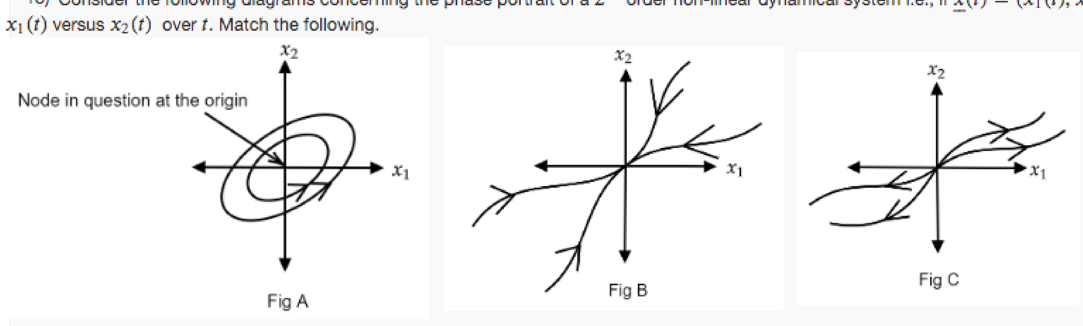
12) How many asymptotic stable fixed points exist towards the solution in question 11?

No, the answer is incorrect.
Score: 0

Accepted Answers:
*(Type: Numeric) 1*

***1 point***

13) Consider the following diagrams concerning the phase portrait of a 2nd order non-linear dynamical system i.e., if $\underline{x}(t) = (x_1(t), x_2(t))^{\mathrm{T}}$ we plot $x_1(t)$ versus $x_2(t)$ over $t$. Match the following. ***2 points***




- Limit cycle: Figure A, unstable: Figure C, stable: Figure B.
- Unstable: Figure B, limit cycle: Figure C, stable: Figure A.
- Stable: Figure C, unstable: Figure B, limit cycle: Figure A.
- Limit cycle: Figure B, stable: Figure A, unstable: Figure C.

No, the answer is incorrect.
Score: 0

Accepted Answers:
*Limit cycle: Figure A, unstable: Figure C, stable: Figure B.*

14) Suppose $\underline{s}$ is a sample vector and $\underline{y}$ is a sample vector of random realization of according to the additional model $\underline{y} = \underline{s} + \underline{w}$. Let $\underline{w} \sim \mathcal{N}\left(\underline{0}, \sigma^2 I\right)$ and $\|\underline{s}\|_2 = 1$. Define $R = \mathbb{E}\left(\underline{y}(n) - \underline{y}(n)^{\mathrm{T}}\right)$. What is the largest eigenvalue of $R$? ***1 point***

- $1 + \sigma^2$
- $1 - \sigma^2$
- $1 + \sigma$
- $1 - \sigma$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$1 + \sigma^2$