# Quiz 5: CS 215, Fall 2024

Prof. Sunita Sarawagi

3 October 2024, 08:30-08:50AM

The quiz will last for 20 minutes. A total of 18 points may be scored. There is no partial marking on questions marked multiple-options-correct.

## 1 Valid Kernel (+2, -1)

Which of the following functions are valid Kernels?

1. $K(x) = sin(x)$
2. $K(x) = \begin{cases} 2 & \text{if } |x| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$
3. $K(x) = \frac{1}{2}e^{-|x|}$
4. $K(x) = \frac{1}{\pi(1+x^2)}$
5. $K(x) = \frac{3}{4}max\{1 - x^2, 0\}$

**Correct Answers: 3, 4, 5**

## 2 Miscellaneous (+3, -2)

Which of the following statements are true?

1. The optimal bandwidth (h) for Kernel Density Estimation is of the order of $O(n^{-\frac{4}{5}})$ (where $n$ is the number of data points).
2. The optimal bandwidth (h, in this case bin-width) for density estimation using histograms is of the order of $O(n^{-\frac{1}{3}})$ (where $n$ is the number of data points).
3. For the optimal bandwidth, bias is minimized.
4. The mean squared error is given by Variance$^2$ + Bias.

**Correct Answer: 2**

## 3 Bias and variance (+3, -1)

State which of the following are true about the histogram and/or KDE density estimator for a distribution $f(x)$. Assume the histogram width is $h$

1. The bias of the histogram estimate when $f(x) = U(0.1, 1)$ and histogram creates bins of width $0.2$ between $0$ and $1$ is zero everywhere.
2. The bias of the KDE estimate reduces with increasing size of data sample.
3. The risk of the KDE reduces faster than histogram's with increasing data size.
4. If $f(x)$ is a Gaussian distribution, then the KDE density with Gaussian Kernel will incur a bias of zero.

**Correct Answer: 3**

## 4 Multivariate Kernel Density Estimation (+3, -1)

Suppose that the datapoints $x_i \in R^d$ (d-dimensional datapoints). The kernel density estimator $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^p}K(\frac{x-x_i}{h})$ (h = bandwidth). Which of the following statements are true?

1. The optimal bandwidth (h) for Kernel Density Estimation is of the order of $O(n^{-\frac{1}{d+2}})$.
2. The mean squared error is of the order of $O(n^{-\frac{4}{d+4}})$.
3. The bias is of the order of $O(h^3)$.
4. The Variance is of the order of $O((nh^d)^{-1})$.

**Correct Answer: 2,4**

## 5 Example Kernel (+2, -1)

Consider the following kernel: $K(x) = N(1-4x^2)$ (defined for the range $|x| \leq \frac{1}{2}$ otherwise 0) (N is the normalization constant). We have four datapoints $x_1 = 0.1, x_2 = 0.4, x_3 = 0.3, x_4 = -0.3$. The bandwidth is $h = 0.3$. What is the value of the kernel density estimator at $x = 0.2$ (Correct to two decimal places)?

**Correct Answer:** 1.39

## 6 Image Segmentation (+3, -1)

The mean-shift algorithm is often used in image segmentation and is based on Kernel Density Estimation (KDE). For this question we will use a Gaussian ($N(0;\sigma)$) Kernel to compute the density function. The goal of the algorithm is to find the closest mode (peak) of the probability density function (pdf) estimated from the data samples $\{x_i\}$ for each sample $x_i$. Essentially, a group of points will converge to the same mode after several iterations, and these points will form a cluster. Unlike K-means, where you need to define the number of clusters, mean-shift automatically determines the number of clusters. Which of the following statements are true about the mean-shift algorithm?

1. We move in the direction of gradient descent of the density function i.e. we move in the direction opposite to the gradient of the density function.
2. We should do the following iteratively: $x = x + \epsilon\nabla f(x)$ where f(x) = Kernel Density Estimator Function
3. $\hat{p}(x) = \frac{1}{N}\sum_{i=1}^{N}K(x-x_i;\sigma)$ where $K(x-x_i;\sigma) = \frac{1}{(2\pi)^d\sigma^d}exp(-\frac{\|x-x_i\|^2}{2\sigma^2})$ where $d$ = dimension of the data and $\hat{p}(x)$ = kernel density estimator.
4. $\nabla\hat{p}(x) = \frac{1}{\sigma^2}\hat{p}(x)m(x;\sigma)$ where $m(x;\sigma) = \frac{\sum_{i=1}^{N}x_i exp(-\frac{\|x-x_i\|^2}{2\sigma^2})}{\sum_{i=1}^{N}exp(-\frac{\|x-x_i\|^2}{2\sigma^2})} - x$

**Correct Answer: 2, 4**

## 7 CDF (+2, -1)

Consider an empirical estimate $\hat{F}(x)$ of the CDF of 10 random variables whose actual CDF is $F(x)$. The highest expected risk of the estimate is

**Correct Answer: 0.025**

**Explanation: The variance of the estimate is $F(x)(1-F(x))/10$. This has a maximum value of 0.25/10**