CS455 Computer Vision: Practice Exam

University of Washington

Instructions:

1. This practice examination contains 23 pages, including this page.
2. For **Multiple-Choice** and **True - False** sections, correct answers will get the points listed next to the question. You will receive 0 for all unanswered questions. You can choose to not answer questions that you are unsure of to avoid receiving a negative point.
3. For **Short Answers**, you will NOT get negative points for incorrect answers. Partial credit will be assigned for showing your work and reasoning.
4. During the Final Exam, you may use one double-sided 8.5” $\times$ 11” **hand-written** sheet with notes that you have prepared. You may not use any other resources, including lecture notes, books, other students or other engineers. These notes must be submitted along with your booklet.

True - False

1. (1 point.) Harris corner detection is always invariant to translation and rotation.

   ◯ True √ **False**

2. (1 point.) The middle layers of a convolutional neural network will learn to detect features such as edges.

   ◯ True √ **False**

3. (1 point.) The Gaussian, Sobel, and Derivative of Gaussian filters are all linearly separable.

   √ **True** ◯ False

4. (1 point.) Even if you don't normalize your data to have zero mean, the first principal component found via PCA will still give you the direction along which projections have the most variance.

   ◯ True √ **False**

5. (1 point.) A derivative filter always sums to 0.

   √ **True** ◯ False

6. (1 point.) Homogeneous coordinates were introduced so that we can encode rotation and scaling into a single matrix multiplication.

   ◯ True √ **False**

7. (1 point.) A 2D translation by a and b in the x and y directions, respectively, has the transformation matrix: $T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & 1 \end{bmatrix}$.

   ◯ True √ **False**

8. (1 point.) In figure 1, the image matrix on the right has more distinct non-zero singular values than the one on the left.

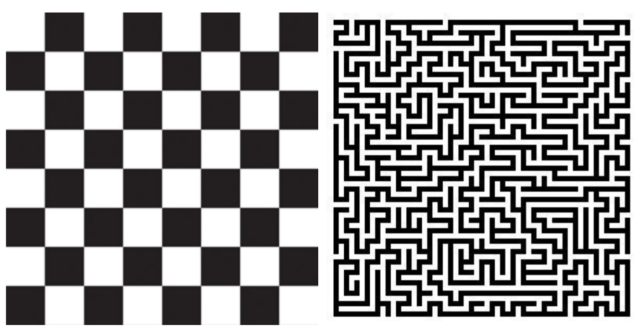
   

   Figure 1: two images with the same width and height.

   √ **True** ◯ False

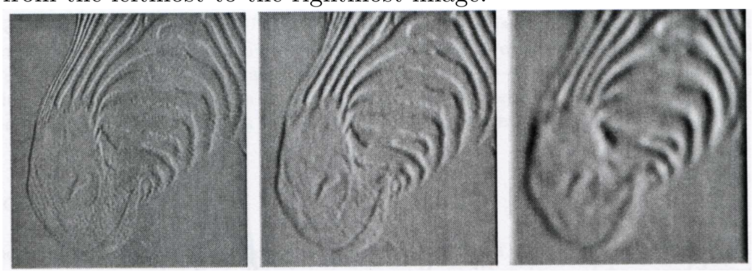
9. (1 point.) Cross-correlation with a 3×1 derivative filter will result in the same output as a convolution with a transposed version of that filter.

   ◯ True √ **False**

10. (1 point.) Cross-correlation is a linear operation.

   √ **True** ◯ False

11. (1 point.) In the figure below, we applied a Gaussian filter to an image of a zebra's nose and computed the derivative. To obtain each of the images, we varied the standard deviation σ of the Gaussian filter. **True or False:** Given a fixed kernel size, The standard deviation σ of the Gaussian filter decreases from the leftmost to the rightmost image.



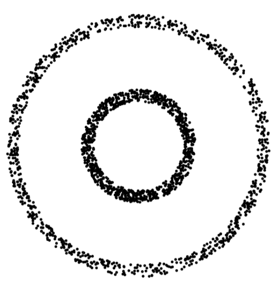
   ◯ True √ **False**

12. √ **True** ◯ False (1 point.) The Histogram of Gradients (HoG) descriptor used in SIFT is invariant to changes in contrast AND changes in exposure.

13. (1 point.) Lucas-Kanade method assumes that the optical flow is constant in a local neighbourhood of the pixel under consideration, and solves the basic optical flow equations for all the pixels in that neighbourhood, by the least squares criterion.

   √ **True** ◯ False

14. (1 point.) Given two rings of points where the rings have equal centers and equal average point values, k-means with Euclidean distances and $k = 2$ will never be able to separate the rings into two clusters, each containing one ring.



   √ **True** ◯ False

15. (1 point.) When using Harris detector, an image region is considered to be a corner if both λ_1 and λ_2 are large. (λ_1 and λ_2 are the eigenvalues of the second moment matrix.)

   √ **True** ◯ False

16. (1 point.) In Horn-Schunk method for optical flow, if the regularization term α is smaller, we will get a smoother flow.

   ◯ True √ **False**

17. (1 point.) The best way to enlarge an image using seam carving is to iteratively find the optimal seam and duplicate it until the image becomes the desired size.

    ◯ True √ **False**

18. (1 point.) If you compute optical flow on a corner feature in a video sequence, a zero optical flow vector can occur despite real-world movement of the object depicted by the video.

    √ **True** ◯ False

Multiple choice answers

1. **RANSAC (2 point).** We use RANSAC to fit a line to a set of points and obtain the fitted line as shown in figure 2(a) below. The dashed line shows all the data points that are considered as inliers. Now we have a dataset with random outliers (the white data points). Using the same set of parameters, RANSAC gives a fitted line as in figure 2(b) below. Of all the listed approaches, which one is most likely to make RANSAC still give a similar result as in (a) on the noisy dataset in (b)? (*Choose the correct answer*):

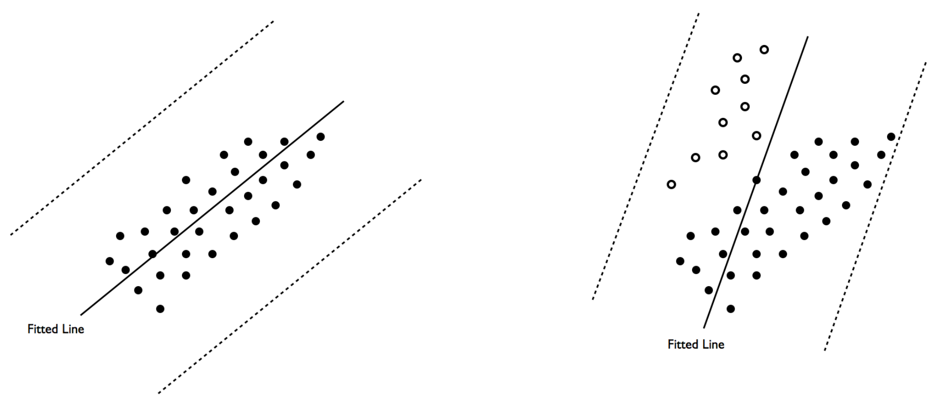


Figure 2: **Left:** Original dataset. **Right:** Noisy dataset.

   ◯ Decrease the number of RANSAC iterations.

   ◯ Increase the number of RANSAC iterations.

   √ **Decrease the inlier error threshold.**

   ◯ Increase the inlier error threshold.

2. **Nearest neighbors (2 point).** Which of the following statements about k-nearest neighbors (k-NN) are true? (*Choose all that apply*):

   ☐ The test classification accuracy is better with larger values of k.

   √ **The decision boundary is smoother with larger values of k.**

   ☐ If k is too large, the classifier is sensitive to noisy points.

   ☐ If k is too large, it is likely that the model will overfit.

3. **RANSAC vs Least squares (2 point).** What are the benefits of RANSAC compared to the least squares method? *Circle all that apply.*

   ☐ RANSAC is faster to compute in all cases

   ☐ RANSAC has a closed form solution

   √ **RANSAC is more robust to outliers**

   ☐ RANSAC handles measurements with small Gaussian noise better

4. **SIFT (2 point).** What does each entry in a 128-dimensional SIFT feature vector represent? (Choose the best answer)

   ☐ A principal component.

   ☐ A sum of pixel values over a small image patch.

   √ **One bin in a histogram of gradient directions.**

☐ The number of pixels whose gradient magnitude falls within some range.

5. **SIFT pipeline (2 points).** Which of the following is a part of the SIFT calculation pipeline? (*Choose all that apply*):

   √ **Difference of Gaussians**

   ☐ Laplacian of Gaussians

   √ **Image Derivatives**

   √ **Histogram of Gradients**

6. **SIFT robustness (2 points).** What is the SIFT feature detector least robust against? (*Choose the correct answer*):

   √ **Occlusion**

   ◯ Rotation

   ◯ Scale

   ◯ Affine illumination

7. **Canny edge detector (2 point).** If a point's neighbors are also above the relevant threshold, the Canny edge detector performs non-maximum suppression among the point and (*Choose the correct answer*):

   ◯ Its 8 neighboring pixels.

   ◯ Its 4 (horizontal and vertical) neighbor pixels.

   ◯ Its neighbors perpendicular to the gradient direction.

   √ **Its neighbors in the gradient direction.**

8. **Rotational invariance (2 points).** Which of the following representations of an image region are rotation invariant? (*Choose all that apply*):

   ☐ A HOG template model

   ☐ Deformable Parts Model

   ☐ Spatial Pyramid Model with SIFT features

   √ **Bag of words model with SIFT features**

9. **Image transformations (2 point).** What processes can be accomplished by expressing an image in homogenous coordinates and applying a single affine transformation matrix? (*Choose all that apply*):

   √ **Resizing an image.**

   √ **Skewing the image.**

   √ **Rotating the image.**

   √ **Translating the image.**

10. **Gaussian filter (2 points).** Consider a 1-D Gaussian with standard deviation of $\sigma = 3$ pixels. Which of the following is the *most efficient* kernel that could capture at least 99% of this distribution? (Hint: 99% of any normal distribution is captured within the interval $(\mu - 3\sigma,\ \mu + 3\sigma)$). (*Choose the correct answer*):

    ◯ Filter kernel of size 1×3

    ◯ Filter kernel of size 1×13

    √ **Filter kernel of size** 1×19

    ◯ Filter kernel of size 1×23

11. **Normalized cross correlation (2 point).** Which of the following are true about Normalized Cross Correlation? (*Choose all that apply*):

    ☐ It is invariant to image rotation.

    √ **It is invariant to image translation.**

    ☐ It is scale invariant.

    ☐ It is commutative.

12. **Deep learning (2 point).** Although your deep learning algorithm will automatically learn the appropriate weights through gradient descent, you as the engineer have to make the following choice(s) about your system (*Choose all that apply*):

    √ **The choice of network architecture.**

    √ **The formulation of the problem (i.e. the output shape and the loss function).**

    ☐ The value for the weights of the first layer.

    √ **The values of the hyperparameters.**

13. **Clustering (2 point).** Which of the following statement is true for k-means and HAC clustering algorithm? (Choose the best answer)

    ☐ k-means and HAC are both sensitive to cluster center initialization.

    √ **k-means and HAC can lead to different clusters when applied to the same data.**

    ☐ k-means works well for non-spherical clusters.

    ☐ HAC starts with all examples in the same cluster, then each cluster is split until we have the desired number of predetermined clusters.

14. **Video Tracking (2 point).** What kinds of image regions are hard to track? *(Circle all that apply)*:

    √ **low texture regions like sky**

    √ **Edges of objects**

    ☐ Corners of objects

    √ **Rotating spheres**

15. **Similarity transformation (2 points).** Which of the following always hold(s) under an similarity transformation? *Circle all that apply.*

    √ **Parallel lines will remain parallel.**

    √ **The ratio between the two areas or two polygons will remain the same.**

    √ **Perpendicular lines will remain perpendicular.**

    √ **The angle between two line segments will remain the same.**

16. **Scale invariance (2 points).** Which of the following representations of an image region are scale invariant? *Circle all that apply.*

    √ **Bag of words model with SIFT features**

    √ **Spatial Pyramid Model with SIFT features**

    ☐ A HOG template model

    ☐ Deformable Parts Model

17. **Optical flow (2 point).** Optical flow is problematic in which conditions? *Circle all that apply.*

    √ **In homogeneous image areas.**

    ☐ In textured image areas.

    √ **At image edges.**

    √ **At the boundaries of moving objects.**

Short Answers 1: Filters, edges, corners, keypoints and descriptors

1. **Filters (4 points).** Give ANY 3x3 example of each of the following types of image **convolution** filters:

   a. **(1 point.)** A darkening filter (only decreases the image intensity)

   b. **(1 point.)** A smoothing filter

   c. **(1 point.)** A left shift filter

   d. **(1 point.)** A filter for detecting horizontal edges

2. **RANSAC (11 points).** RANSAC is a powerful method for a wide range of model fitting problems as it is easy to implement. In class, we've seen how RANSAC can be applied to fitting lines. However, RANSAC can handle more complicated fitting problems as well, such as fitting circles. In this problem we will solve for the steps needed to fit circles with RANSAC.

   (a) **(2 point).** What is the minimum number of points we must sample in a seed group to compute an estimate for a uniquely defined circle?

   (b) **(5 points).** If we obtain a good solution that has few outliers, we want to refit the circle using all of the inliers, and not just the seed group. For speed purposes, we would like to keep the same center of the estimated circle from the seed group, and simply refit the radius of the circle. The equation for our circle is given as:

$$(x - x_c)^2 + (y - y_c)^2 = r^2$$

   where (x_c, y_c) are the coordinates of the center of the circle, and r is the radius. The error we would like to minimize is given by:

$$E(x_c, y_c, r) = \sum_{i=1}^{m} \left(\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} - r\right)^2$$

   where m is the number of inliers. Derive the new radius r that minimizes this least squares error function.

(c) **(4 points).** One of the benefits to RANSAC is that we are able to calculate the failure rate for a given number of samples. Suppose we know that 60% of our data is outliers. How many times do we need to sample to assure with probability 20% that we have at least one sample being all inliers? You can leave your answer in terms of log functions.

Assume that there are n data points and present your answer in terms of n.

3. **Convolution (5 points).** Recall that the definition of a 1D convolution between two functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ is a function $f * g : \mathbb{R} \to \mathbb{R}$ defined by:

$$(f * g)[x] = \sum_i f[i]g[x - i]$$

Prove that 1D convolution is **associative**. In other terms, given three functions f, g and h, show that:

$$f * (g * h) = (f * g) * h$$

Hint: you can define $a = f * (g * h)$ and $b = (f * g) * h$ and show that $a = b$.

Solution:

$$\begin{aligned}
a(x) &= (f * (g * h))[x] \\
&= \sum_i f(i)(g * h)[x - i] \\
&= \sum_i f(i) \sum_j g(j)h(x - i - j) \\
&= \sum_i f(i) \sum_k g(k - i)h(x - k) && \text{where } k = i + j \\
&= \sum_k h(x - k) \sum_i f(i)g(k - i) \\
&= \sum_k h(x - k)(f * g)(k) \\
&= ((f * g) * h)[x] \\
&= b(x)
\end{aligned}$$

4. **Hough Transform (5 points).** Your aim is to detect lines in an image using Hough transform. Using Canny edge detector, you have found a set of 5 points that correspond to edges in the first figure below. Using the Hough Transform, transform each point into a line from x, y space to a, b space. Draw these 5 lines in the second image, and find the best values of a, b that define the line. Note that a line is defined as $y = ax + b$.

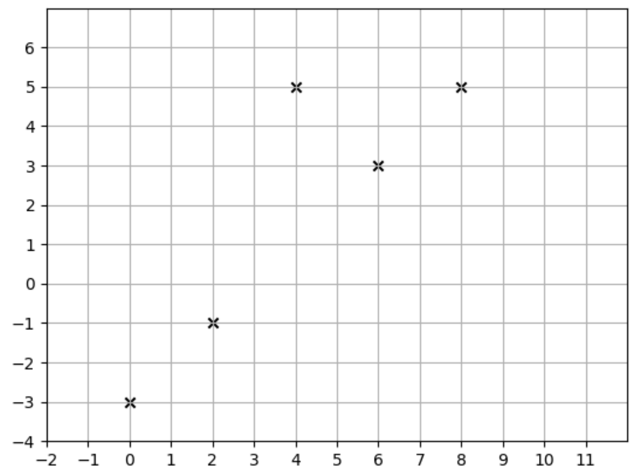


Figure 3: x (horizontal), y (vertical) space where we have the detected points

images/hough_ab_solution.png

Figure 4: a, b space

Best value of $a =$ ___**1**___. Best value of $b =$ ___**-3**___.

Short Answers 2: Segmentation and seam carving

1. **k-means local minima (5 points.)** Suppose we want to organize m points $x_1, ..., x_m$ into k clusters. This is equivalent to giving a label $l_1, ..., l_m$ to each point, where $l_i \in 1, ..., k$. The k-means clustering algorithm solves this problem by minimizing the quantity:

$$\sum_{i=1}^{m} d(x_i, \mu_{l_i}),$$

where μ_c is the centroid of all points x_i with label $l_i = c$ and $d(p, q)$ is the Euclidean distance between points p and q.

Consider points $x_1 = (0, 16), x_2 = (0, 9), x_3 = (-4, 0), x_4 = (4, 0)$. Suppose we wish to separate these points into two clusters and we initialize the k-means algorithm by randomly assigning to clusters. This random initialization assigns x_1 to cluster 1 and all the other points to cluster 2.

(a) **(3 points).** With these initial assignments of points to clusters, what will be the final assignment of points to clusters computed by k-means? Explicitly write out the steps that the k-means algorithm will take to find the final cluster assignments.

Solution: x_1 is assigned to cluster 1 and the rest are assigned to cluster 2.

(b) **(2 point).** In this case, does k-means find the assignments of points to clusters that minimize the quantity? What is the optimal cluster assignment?

Solution: The optimal solution: x_1 and x_2 belongs to cluster 1 and x_3 and x_4 belongs to cluster 2.

2. **Harris Corner Detector (5 points).** Consider the case of running Harris Corner Detection on a patch in this chessboard image. The white squares have value 2 and the black squares have value 0. You can assume a square window function ($w(x, y) = 1$ everywhere in the image patch), and that the patch used to compute the second moment matrix is exactly 2×2 pixels.

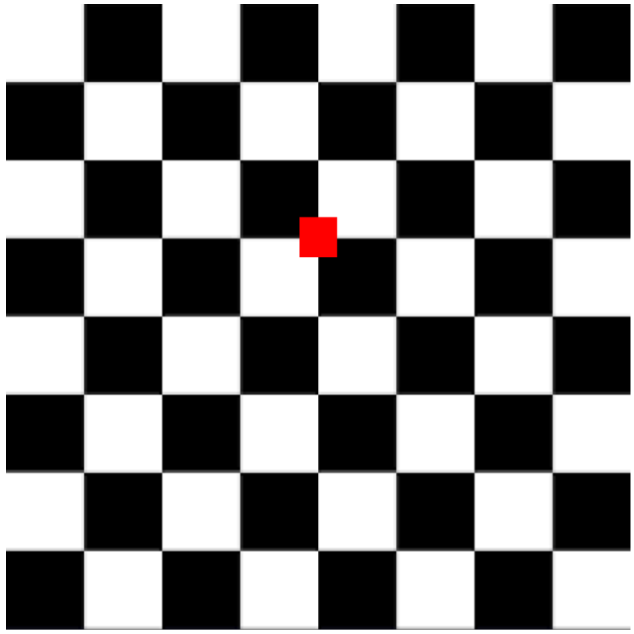


Figure 5: Checkerboard with a patch of size $(2, 2)$ pixels

Specifically, the patch matrix looks like this:

$$P = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$$

For taking image derivatives, we can **zero-pad** and use these standard filters:

$$D_x = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$$
$$D_y = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T$$

(a) **(2 points).** What is the second moment matrix?

Solution: First, we can compute the image derivatives:

$$I_x = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$
$$I_y = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

And then the second moment matrix:

$$\begin{aligned} M &= \sum_{x,y} w(x,y) \begin{bmatrix} I_x(x,y)^2 & I_x(x,y)I_y(x,y) \\ I_x(x,y)I_y(x,y) & I_y(x,y)^2 \end{bmatrix} \\ &= \begin{bmatrix} 2^2+2^2 & 2*2+2*2 \\ 2*2+2*2 & 2^2+2^2 \end{bmatrix} \\ &= \begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix} \end{aligned}$$

(b) **(1 points).** What are the eigenvalues of the second moment matrix?

Solution: Finding the eigenvalues can be done by inspection:

$$\lambda_1 = 0$$
$$\lambda_2 = 16$$

(c) **(1 point).** The patch is classified as a corner.

◯ True √ **False**

(d) **(1 point).** Would the classification change if we increased the size of the patch? Why or why not?

Solution: In the 2×2 case, the "corner" patch looks identical to a diagonal edge. Using a larger patch would result in correct classification as a corner.

3. **Seam Carving (5 points.)**

(a) **(1 points).** Impressed by the results of seam carving for content-aware image re-sizing, you decide to play around with different combinations of scaling and re-sizing through seam carving. One particular combination involves first downsizing an image using standard scaling and then applying seam carving techniques to enlarge the image back to its original size. What is the effect of doing this on the image?

> **Solution:** Contents of image are downsized, low energy parts (for example, background) are enlarged.

(b) **(2 points).** During the seam carving lecture, we discussed some issues with the standard energy cost and the need for the forward energy cost. Describe the issue with using the standard energy cost and provide an example of when these effects are most noticeable. How does the forward energy cost function account for this issue?

> **Solution:** The standard energy cost ignores energy that is inserted into the new image. Standard energy cost fails when image does not contain 'less important' areas or where the content layout prevents seams from avoiding important parts.
>
> Forward energy removes the seam whose removal inserts the minimal amount of energy into the image.

(c) **Seam Carving with Face Detector (2 points).** We learned about object detection and even implemented our own face detector. Recall that the original, unconstrained seam carving method could potentially result in distorted faces. How would you make use of a face detector, to ensure that faces remain unaltered during image re-sizing?

Solution: Since seam carving removes seams with low cost, use the face detector to find areas with high face response. Set the cost at these areas to high values.

Short Answers 3: Classification and detection

1. **Classification concepts (4 points).**

   (a) **(1 points).** When building a classifier, you find that your model performs really well on the training set but fails to do well on the test set. Is the model overfitting or underfitting? Does the model have high or low bias and high or low variance?

(b) **(2 point).** In LDA, we maximize the "between class scatter" while minimizing the "within class scatter". To save compute, your homework partner decides to only maximize the "between class scatter" but does not place any restrictions on the "within class scatter". How will this impact the accuracy of a classifier that is using the the LDA features? What can you say about the clusters formed by the data points in the LDA feature space?

(c) **(1 point).** How does the stride in "sliding window" based detection models impact the accuracy of the models? How does it impact runtime?

2. **SVD and Eigendecomposition (5 points)** Consider a $m \times n$ real matrix M. We apply SVD to M and obtain a decomposition $M = U\Sigma V^T$. Σ is a diagonal matrix of shape (m, n), and the values on its diagonal are called the singular values $\sigma_i, i \in [1, min(m, n)]$. U and V are orthogonal matrices, which means that $U^T U = I$ and $V^T V = I$.

   Now consider MM^T. This matrix is symmetric, and has real eigenvalues $\lambda_i, i \in [1, m]$.

   Prove that if we order the λ_i in decreasing order and the σ_i in decreasing order, we have:

$$\forall i \in [1, min(m, n)], \quad \lambda_i = \sigma_i^2$$

Solution: We can write MM^T using the SVD:

$$\begin{aligned} MM^T &= U\Sigma V^T (U\Sigma V^T)^T \\ &= U\Sigma V^T V \Sigma^T U^T \\ &= U\Sigma\Sigma^T U^T \end{aligned}$$

We can see that the columns of U are the eigenvectors of MM^T:

$$\begin{aligned} MM^T U &= U\Sigma\Sigma^T U^T U \\ &= U\Sigma\Sigma^T \end{aligned}$$

If we consider the equality column by column, we obtain:

$$\forall i \in [1, m], \quad MM^T u_i = \sigma_i^2 u_i$$

where u_i is the i^{th} column of U.
Therefore, the u_i are eigenvectors of MM^T associated with eigenvalues σ_i^2. Therefore, we have $\sigma_i^2 = \lambda_i$.

3. **Nearest Neighbors (2 points).** Consider a dataset with C classes, N training examples, and T test examples. The number of class examples for the train and test set, respectively, are stored in numpy arrays $n = [n_1, \ldots, n_C]$ and $t = [t_1, \ldots, t_C]$. Here n_j represents the number of training examples in class j, and t_j represents the number of test examples in class j. What is the test accuracy of a k-Nearest Neighbor classifier with $k = N$?

Solution: Everything will be classified in the biggest training class $np.argmax(n)$ so the test accuracy will be $t[np.argmax(n)]/T$

4. **kNN Boundaries (2 points).** Figure 6 shows a training set of 14 examples, with 7 example for class 1 (with the crosses) and 7 examples for class 2 (with the circles).

   If we apply k-nearest neighbors, we can classify any testing point into class 1 or class 2. If we do that for every point in $\mathbb{R}^2$, we obtain decision boundaries which delimit zones of $\mathbb{R}^2$ where the classification is either 1 or 2.

   Draw on figure 6 the decision boundaries for the k-nearest neighbors algorithm with $k = 1$. **Use L_2 distance.**

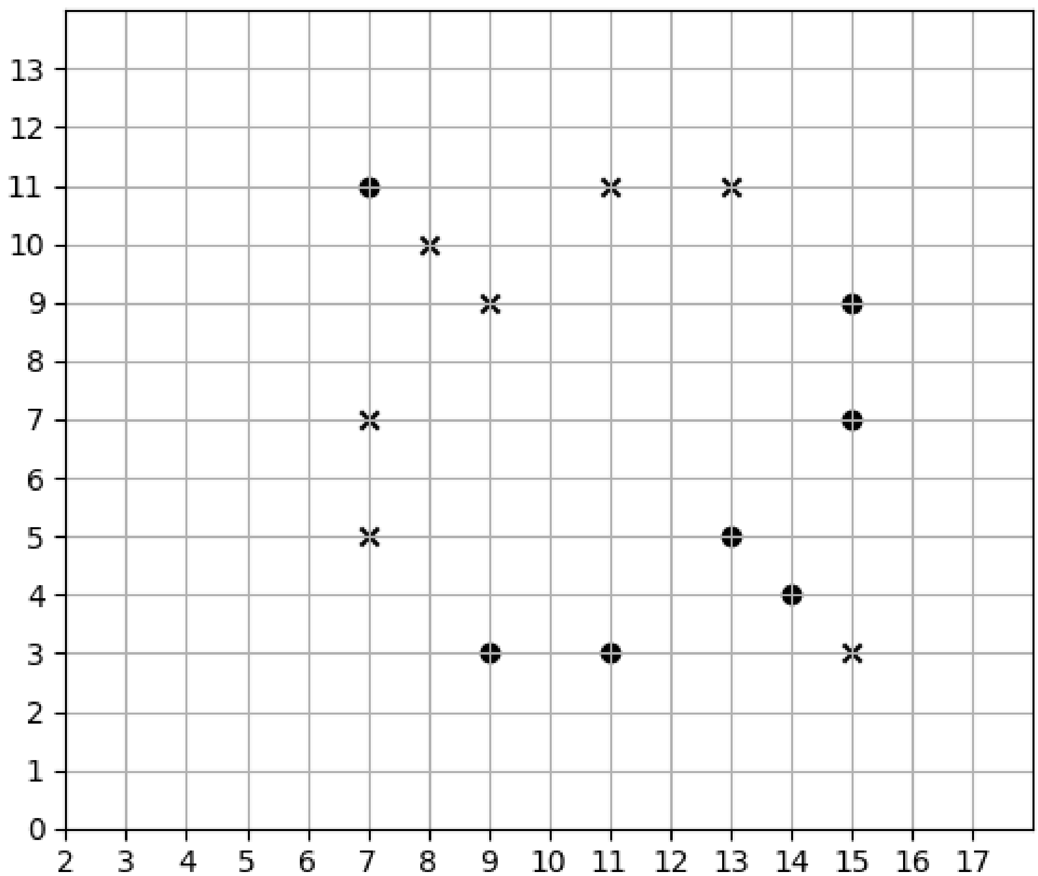


Figure 6: Decision boundaries for kNN with $k = 1$ and L_2 distance