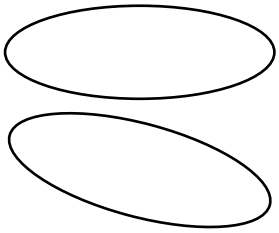
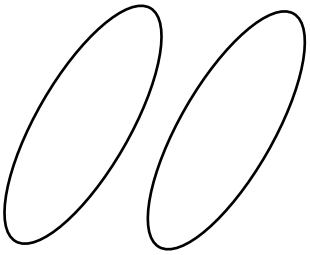
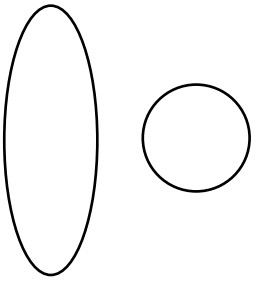
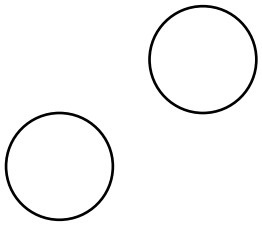
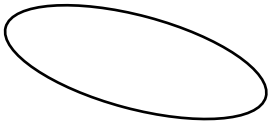
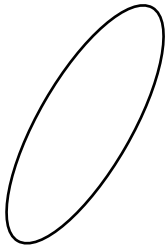
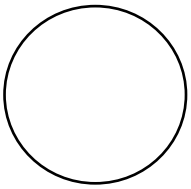
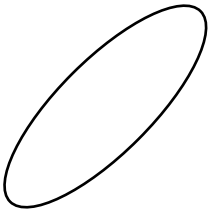
1	2	3	4	5	6	Σ

……

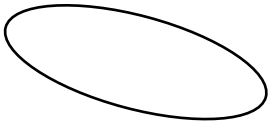
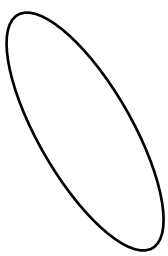
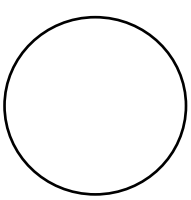
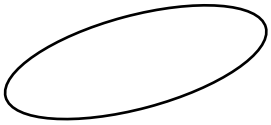
1. (2 p) Which of the following pairs of classes have linear decision boundary between classes? Check all appropriate class pairs.

a. ☐	b. ☐	c. ☐	d. ☐
			

2. (2 p) For which of the following classes given by their contour diagrams a > b holds, where a and b are elements of the covariance matrix: $\begin{bmatrix} a & c \\ c & b \end{bmatrix}$? we assume that coefficient *a* is computed for the feature represented on the horizontal axis, and coefficient *b* for the feature represented on the vertical axis. Check all appropriate classes.

a. ☐	b. ☐	c. ☐	d. ☐
			

3. (2 p) Which of the following classes given by their contour diagrams have negative correlation between features, i.e. c < 0 in the covariance matrix: $\begin{bmatrix} a & c \\ c & b \end{bmatrix}$?. we assume that coefficient a is computed for the feature represented on the horizontal axis, and coefficient b for the feature represented on the vertical axis. Check all appropriate classes.

a. ☐	b. ☐	c. ☐	d. ☐
			

4. (9 p) We plan to implement bottom-up clustering with single linkage (distance between two clusters is the minimum distance between pairs of points, in which one point belongs to the first and the other to the second cluster). The number of points (N) is moderate so we can afford to store cluster distance matrix.

(2p) Propose memory efficient structure to store distances between clusters. (*How many distances are there?*)

(3p) Which elements of the distance matrix will be invalidated by a join? (*How many distances should be deleted? How many distances should be updated?*)

(4p) Propose update procedure for distance matrix which will minimize number of computations needed.

5. (9 p) I prepared 3-NN classifier with the recognition quality slightly better that my customer requirements. The problem is too high number of samples in the training set (both in memory and classification time). I plan to use edition and reduction of the training set based on proximity graph, but I don't want to change classification result.
   (3p) Why can't I use directly algorithm version presented on the lecture?

   (6p) Propose changes in these algorithms making it possible to use them in my problem.

6. (7 p) Order classifiers listed below according to classification speed (1 – fastest, 3 – slowest) and show approximate number of operations needed to classify one unknown sample. Assume that no special acceleration techniques are used but all parameters that can be computed off-line are in fact precomputed.

   We assume N=105 samples in the training set, D=4 dimensions of the feature space and C=7 classes.

   …….. **3-NN classifier**

   ……

   …….. **naive Bayes classifier (assuming independent features)**

   ……

   …….. **two vs. rest linear classifier ensemble (Attention: two vs. rest!!!)**

   ……