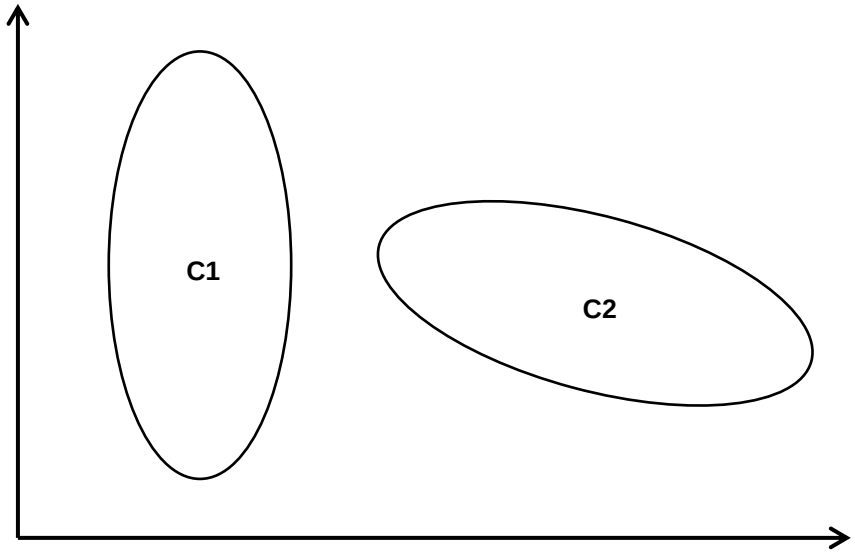
1	2	3	4	Σ

……

1. (6 pts) We have two classes with normal probability distributions. Their contour plots are shown on the diagram below.



   a. (2 pts) The covariance matrices of both classes are in the form $\begin{bmatrix} a & c \\ c & b \end{bmatrix}$ (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). Which of the following statements are true for class $\mathbf{C_2}$ (check all that apply):

   ☐ a == b ☐ a > b ☐ a < b ☐ c > 0 ☐ c < 0 ☐ c == 0

   b. (2 pts) Sketch the decision boundary for equal a priori probabilities $P(C_1) == P(C_2)$ and the same loss values for different errors.

   c. (2 pts) Sketch the decision boundary for a priori probabilities $P(C_1) > P(C_2)$ and the same loss values for different errors.

2. (7 pts) For the classification of C = 7 classes, we use an ensemble of one versus one linear classifiers. Preparation of the ensemble consists in training an elementary linear classifier for each pair of classes. We use a classification scheme in which ensemble indicates a particular class as a classification result if all elementary classifiers which during the training "saw" the class give a consistent indication of this class (just one other decision of elementary classifier makes the ensemble to output reject decision). With such a classification scheme we achieved *Rec* recognition coefficient, *Err* error coefficient and *Rej* reject decision coefficient.
   Of course *Rec* + *Err* + *Rej* = 1.

   a. (1 pt) How many classifiers are there in the ensemble?

   b. (1 pt) What is the maximum number of votes that can fall on any class?

   c. (3 pts) How will these coefficients change if, instead of unanimity voting as above, we use a simple majority vote?

   d. (2 pts) Is finding a class with maximum number of votes enough to prepare unbiased classifier (i.e. classifier without a systematic error)?

3. (9 pts) We plan to implement bottom-up clustering using maximum cluster distance with Euclidean distance between points. Training set (i.e. individual samples coordinates) are stored in rows of *TSet* matrix. Column vector *CLab* (cluster label) contains for each point in the training set cluster index.

   a. (3 pts) Write pseudocode to select two clusters to merge.

   b. (2 pts) What is computational complexity of selecting two clusters to merge?

   c. (2 pts) Which of results can be reused in the next iteration?

   d. (2 pts) What memory is needed to store these reusable results?

4. (9 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=150 samples in the training set, D=4 dimensions of the feature space and C=5 classes.

   …….. **multivariate Bayes classifier**

   ……

   …….. **1-NN classifier**

   ……

   …….. **one vs. rest linear classifier ensemble**

   ……