Implementation:

1. Select the dataset MINST from <http://yann.lecun.com/exdb/mnist/> and merge the training and testing datasets into one single one big dataset.
2. Using the method described in the attached paper (section 3.2), generate a dataset for anomaly detection tasks. That is, some classes will be merged into normal classes and some classes will be merged into abnormal classes. From the abnormal classes, down sample them to 5%.
3. Select three benchmark approaches often used for anomaly detection tasks: two classical approaches and one based on deep learning.
4. Perform a benchmark analysis among the selected approaches showing results compatibles for tasks of anomaly detection. Do not forget to normalize data, impute missing values (if exist), etc.
5. The trained models should only use normal data samples. In case you decide to split the normal data into training, validation and test, the training and validation sets should ony include normal data whereas the test set should merge both normal and abnormal samples.
6. Prepare a presentation showing (not more than 15 slides):
   1. Dataset characteristics
   2. Results from the dataset generation for anomaly detection tasks
   3. Steps done in the preprocess of data (if any done)
   4. Benchmark methods with explanations (do not forget to cite literature)
   5. Results
   6. Conclusions

You can implement the exercise using matlab or python (maybe it would be better for license issue to use sklearn packages thou and keras)

Theoretical exercises:

1. Assume that you have gathered a set of N multidimensional data points each of which with dimension d.  
   Furthermore, assume you decide to approximate the data distribution using a multivariate Gaussian distribution.  
   What method you would use to estimate the sufficient statistics of this distribution. Write down formally the equations to estimate the necessary parameters.
2. Following the point 1, what if the expectation is assumed to be Gaussian distributed with prior hyperparameters mu0 and sigma0.  
   How does this information change the estimation of the sufficient statistics in the point 1?
3. Suppose you get an additional data sample. How can you test whether the new data point belongs to the estimated distribution in point 1, or it belongs to some other multivariate Gaussian distribution with  
   different expected value but same covariance matrix? Write down formally any equation you believe would help you solving the problem.

If there are questions, try to answer them alone and set up assumptions you may need to make.

Good luck!