**Technische Universität München**
**Fakultät für Informatik**
**Prof. N. Navab, Ph.D.**

**SS 2015**
**Exercise Sheet 3**
**April 28, 2015**

**Practical Course: Machine Learning in Medical Imaging**

# Evaluation

## 1 Confusion Matrix

Given the ground truth and predictions of a classifier for $k$ classes, construct a $k \times k$ confusion matrix that summarizes the classifier's performance. For any number of classes you should be able to calculate the following 6 measures based on a confusion matrix: sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and $F_1$ measure. Keep in mind that you have to provide an option to define the class of interest in order to calculate these values.

### Tasks

1. Write a function `confusion_matrix` which expects two arguments: A vector containing the ground truth for each sample, and a vector containing the predicted class for each sample. The returned value should by a $k \times k$ matrix, where $k$ indicates the number of classes present in the ground truth.
2. Write a function `one_vs_all` which takes the index of the class of interest and a single $k \times k$ matrix as returned by `confusion_matrix`. The function should return a single $2 \times 2$ confusion matrix with respect to the provided class according to the one-vs-all principal. For instance, if the data set has three classes *apple*, *orange*, and *banana*, you should be be able to obtain 3 different confusion matrices.
3. Write six functions – one for each performance measure mentioned above – which expect a $2 \times 2$ confusion matrix as returned by `one_vs_all`. The performance measure should be calculated with respect to the class of interest as defined in the call to `one_vs_all`.
4. Create one $2 \times 2$ confusion matrices for each logistic regression model you trained in the previous exercise.

## 2 ROC and Precision-Recall Curve

Now assume that we can obtain multiple confusion matrices because our *binary* classifier is able to assign each prediction a probability or score. Given a threshold $t$ all predictions with probability smaller than $t$ are classified as negative and positive otherwise. Gradually increasing this threshold from 0 to 1, we can can construct one confusion matrix for each threshold.

### Tasks

1. Create a function `threshold_confusion_matrix` that expects a vector containing the ground truth and a vector of containing the predicted probabilities for each sample. For each **unique** threshold the function should return a $2 \times 2$ confusion matrix. The implementation should iterate over the predictions only a single time, i.e. its complexity should be $O(n)$.
2. Based on the list of confusion matrices obtained by `threshold_confusion_matrix` you can easily derive all the performance measures you already implemented. Construct a ROC and precision-recall curve for the different logistic regression models you created for the *South African Heart Disease* data set. Which model performs best?

# 3 Validation

Previously, we assessed the performance of our models merely on the training error, which results in bad estimates of the classifier's true performance. Hence, we want to train and test the classifier on two disjoint sets.

### Tasks

1. Perform 10-fold **stratified cross-validation** on the `SAheart` data set using logistic regression and for each fold obtain a ROC and precision-recall curve.