Application Background

Computer vision and image analysis are becoming more and more important in sports analytics, the science of analyzing and modeling processes underlying sporting events. Sports with a high media coverage create a demand for systematic review and objective evaluation of the performance of individual athletes as well as of teams. Across almost all sports, management and coaches make use of statistics and categorized video material to support their strategies.

In this assignment, we consider classication of the protagonists in soccer. A pattern detection algorithm has extracted regions of interest (ROIs) from a live video recording. Now each ROI has to be classied into one out of seven main classes. We have the obvious ve classes outelder (two classes, one for each team), goalkeeper (two classes) and referees. Further, there is a category that groups ROIs that contain at least one outelder from each team. Finally, there is a class for irrelevant objects (i.e., false detections by the ROI detection algorithm). Sample ROIs are shown in Figure 1, the classes are summarized with the corresponding class index (label) in Table 1.

Because teams and referees can be identied based on the color of the clothing, color histogram features are extracted from the ROIs. The color model is HSV (hue, saturation, value), considering 3, 3, and 2 bits per channel, respectively.

The Exercises

Question 1

(data understanding and preprocessing). Download and extract the data.
Consider the training data trainInput.csv and the corresponding labels trainTarget.csv.
Plot a histogram showing the probability distribution of classes in the training data.
The ith row in trainInput.csv are the features of the ith training pattern. The class label of the ith pattern is given in the ith row of trainTarget.csv.
Deliverables: description of software used; single histogram plot

Question 2

(principal component analysis). Perform a principal component analysis of the training data trainInput.csv. Plot the eigenspectrum (see Figure 12.4 in [1] for an example). How many components are necessary to \explain 90% of the variance"? Visualize the data by a scatter plot of the data projected on the rst two principal components. Use di
erent colors for the di
erent classes in the plot (see Figure 12.8 in [1] for an example, which, however, lacks proper axes labels).
Deliverables: description of software used; plot of the eigenspectrum; indicate number of components necessary to explain 90% of variance; scatter plot of the data projected on the rst two principal components with di
erent colors indicating the 7 di
erent classes

Question 3

(clustering). Perform 7-means clustering of trainInput.csv (feel free to play around with the number of clusters). After that, project the cluster centers to the rst two principal components of the training data. Then visualize the clusters by adding the cluster centers to the plot from the previous exercise. Brie
y discuss the results: Did you get meaningful clusters?
Deliverables: description of software used; one plot with cluster centers and data points;

short discussion of results

Question 4

(overtting). John Langford, who is \Doctor of Learning at Yahoo Research", maintains a very interesting blog (web log). Read the very true blog entry: \Clever methods of overtting," http://hunch.net/?p=22, 2005. Discuss if and how the di erent types of overtting can occur when applying machine learning techniques to our sample sports analytics application. Ignore the last type of overtting. You need not discuss issues related to reviewing of scientic papers (still, it is good to keep them in mind).

Deliverables: short discussion addressing the rst 10 \methods of overtting" listed in the blog entry

Question 5

(multi-class classication). Use a linear and a non-linear classication method (picking from the methods presented in the course) for classifying the 7 image classes, for example k-nearest neighbor and linear discriminant analysis (LDA). Use trainInput.csv and trainTarget.csv for training. After you trained a model, use the test data in testInput.csv and testTarget.csv to evaluate it. Report the classication error on both training and test set.

Deliverables: description of software used; arguments for your choice of classication methods; a short description of how you proceeded and what training and test results you achieved

Question 6

(binary classication using support vector machines). Now we consider binary classication using support vector machines (SVMs). To this end, we reduce the problem to distinguishing between referees and regions not containing persons of interest. Please use the data les trainInputBinary.csv, trainTargetBinary.csv, testInputBinary.csv, and testTargetBinary.csv. These are real-world data. The splitting into training and testing data has been done by the people providing the images. The class frequencies in training and testing data set di er considerably. This could be bad luck, but could also indicate a sampling bias (i.e., a violation of the i.i.d. assumption). The di erences between training and test set make the problem dicult for some learning algorithms. Please ignore this phenomenon in the assignment, that is, do not use asymmetric loss functions, class-dependent regularization parameters (i.e., di erent\C-values" depending on the class), etc.1

For this exercise, use standard C-SVMs as introduced in the lecture. Employ radial Gaussian kernels of the form

$k(x; z) = \exp(.. kx .. zk2)$:

Here > 0 is a bandwidth parameter that has to be chosen in the model selection process. Note that instead of often the parameter $= p 1=(2)$ is considered.

Jaakkola's heuristic provides a reasonable initial guess for the bandwidth parameter or

of a Gaussian kernel [3]. To estimate a good value for , consider all pairs consisting
of a training input vector from the positive class and a training input vector from the
negative class. Compute the di
erence in input space between all pairs. The median of
these distances can be used as a measure of scale and therefore as a guess for . More
formally, compute
G = fkxi .. xjk j (xi; yi); (xj ; yj) 2 S ^ yi 6= yjg
1Be careful, because, for instance, the SVM from the Matlab Bioinformatics Toolbox may by
default
use di
erent regularization parameters depending on the class and the class frequency.
5
based on your training data S. Then set Jaakkola equal to the median of the values in G:
Jaakkola = median(G)
Compute the bandwidth parameter
Jaakkola from Jaakkola using the identity given above.
Use grid-search to determine appropriate SVM hyperparameters
and C. Look at all
combinations of
C 2 fb..1; 1; b; b2; b3g
and

2

Jaakkola bi j i 2 f..3;..2;..1; 0; 1; 2; 3g

;
where the base b can be chosen to be either 2, the base e of the natural logarithm (Euler's
number), or 10. Feel free to vary this grid. For each pair, estimate the performance of the
SVM using 5-fold cross validation (see section 1.3 in [1]). Pick the hyperparameter pair with
the lowest average 0-1 loss (classication error) and use it for training an SVM with the com-
plete training dataset. Only use trainInputBinary.csv and trainTargetBinary.csv in
the model selection and training process.
Report the values for C and
you found in the models selection process. Compute the
classication accuracy based on the 0-1 loss on the training data as well as on the test
data testInputBinary.csv and testTargetBinary.csv. An accuracy on the test set
signicantly larger than 75% can be expected.
Deliverables: description of software used; a short description of how you proceeded; initial

or value suggested by Jaakkola's heuristic; optimal C and
found by grid search;
classication accuracy on training and test data
Question 7
(linear regression vs. classication I.). 2 In this exercise, we explore the relation
between regression and linear binary classication. Let the input space be X = Rp and
the output space be Y. We are given ` training patterns S = f(x1; y1); : : : ; (x`; y`)g. Let
`1 of these patterns belong to the rst class and `2 to the second class, respectively (i.e.
` = `1 + `2).
Show that the LDA classication rule as introduced in the lecture classies (see also

section 4.3 in [2]) to the second class if
xT^..1
(^2 .. ^1) >
1
2
^T
2
^
..1
^2 ..
1
2
^T
1
^
..1
^1 + ln
`1
`
.. ln
`2
`
and to the rst class otherwise. Here ^
is the empirical covariance matrix and ^1 and
^2 are the means of the training data from the rst and second class, respectively.
Assume that someone uses least squares regression in order to solve a binary classica-
tion problem. Further, assume that the two classes are encoded by Y = f..`=`1; `=`2g.
2This exercise has been taken from [2].
6
Consider the regression objective function
X`
i=1
(yi .. 0 .. Txi)2 ;
with regression coecients 2 Rp and scalar o
set/bias/intercept parameter 0.
Show that the optimal solution ^
minimizing this least squares criterion is in the
direction
^
/ ^
..1
(^2 .. ^1) (/ means proportional) :
Therefore the least squares regression coecient is identical to the LDA coecient,
up to a scalar multiple.3
Hints (feel free to ignore): There are several ways to prove the statements. For the second
part, you may nd it convenient to get rid of 0 as in (3.3) in [1]. You may want to consult
[4] (in particular section 2.4.2) to get the derivatives of vectors and matrices right.
Deliverables: proofs including intermediate steps
Question 8
 (linear regression vs. classication II.). As the previous exercise shows, abusing
regression for binary classication is not a completely crazy thing to do. Does this also

hold for multi-class classication, in which the same least squares regression criterion is used and the class indices serve as the targets for regression (e.g., Y = f0; 1; : : : ; 6g)? You may conduct an experiment to support your answer, but you have to provide arguments beyond your empirical ndings.
Deliverables: reason(s) why this is a reasonable approach or not.

References

[1] C. Bishop. Pattern Recognition and Machine Learning. Springer, 1 edition, 2006.
[2] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. Springer, 2009.
[3] T. Jaakkola, M. Diekhaus, and D. Haussler. Using the Fisher kernel method to detect remote protein homologies. In T. Lengauer, R. Schneider, P. Bork, D. Brutlad, J. Glas-gow, H.-W. Mewes, and R. Zimmer, editors, Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology, pages 149{158. AAAI Press, 1999.
[4] K. B. Petersen and M. S. Pedersen. The matrix cookbook. Technical University of Denmark, 2008.
3Actually, this result holds for any distinct coding of the two classes. However, in general the o
-
set/intercept/bias parameter di
ers depending on whether LDA or least squares regression is used to determine it. For balanced class frequencies the two approaches coincide.