First Group Pick 5 questions, 1 pts per question

Why Cascading

What could have been a good motivation for Viola and Jones to use cascading?

Although true detection was good, cascading could improve it further

False positive rate (FPR) was high, so cascading could help reduce it

Cascade could reduce the overall FPR by cascading even though overall true detection rate (TDR) would be lower

Cascading is always useful, so they used it

Correct Answer

2, 3

1, 3

2, 4

1, 2

Overfitting

Suppose we have an ensemble of classifiers that have collectively achieved an accuracy of 100% on a training set. This is a classic sign of overfitting and you must begin to implement regularization methods in order to reduce the chance of performing poorly on your test set.

True

Correct Answer

False

Face Recognition in the Wild

Suppose you work at a company that is interested in deploying facial recognition at scale. Your company wants to detect if a picture contains a human face in it and will use the decision to intelligently recommend stylistic filters that a user could apply to the image before posting the image to a social feed. Since detection in this application needs to happen in real-time, you decide to scan a template face across a given image and then compute the similarity between the template and patches of the image. You will mark patches with high similarity to the template image as faces. In doing so, you are making several assumptions about the data your application will see. Please select all of the assumptions:

Correct Answer

All faces will be seen head-on (i.e., every face will be looking directly at the camera).

Correct Answer

All faces will be of the same size as your template face.

The distribution of the dataset (with regards to the samples' age and race) used to compute the template face will have no effect on the types of faces it can recognize.

Every image the application sees will have a face in it.

Performance Curves 1

Near the end of lecture, we provided a modification to our standard boosted classifier, adding in a term to adjust classification: LaTeX: H\_{\textrm{modified}} = \textrm{sign}(\Sigma\_t \alpha\_t h\_t(x) + Y)H modified = sign ( Σ t α t h t ( x ) + Y ). Here, LaTeX: YY is a real-valued parameter. If we continually increase LaTeX: YY so that it is a large positive number, what do you expect to see happen to the performance of the classifier:

N.B.: There are two questions like this one. They test two different situations. Please read the question carefully.

The false rejection rate will increase.

Correct Answer

The false rejection rate will decrease.

Correct Answer

The True Positive Rate (also known as sensitivity or recall) will increase.

The True Positive Rate (also known as sensitivity or recall) will decrease.

Correct Answer

The False Positive Rate will increase.

The False Positive Rate will decrease.

Number of Computations

Suppose you are given an image that is 500 pixels by 800 pixels. Also suppose you are given a face template that is 20 pixels by 20 pixels. We will slide the template face over the image in increments of one pixel. At each increment, we compute a dot product between the patch of image and the template face. How many dot products will we need to compute?

374400

Correct Answer

375661

400000

373141

Viola Jones 1

The rectangular features used in Voila-Jones algorithm are not the best features for learning image representations and better features could be used. Why did the authors decide to still use them in their algorithm? (Hint: it would help to read their paperLinks to an external site.).

Correct Answer

The integral image allowed rectangular features to be computed very efficiently

The authors did not do their literature survey

Rectangular features are the only features that work with AdaBoost

Rectangular features enabled the features to be cascaded

Similarity Metrics

Given a template face (for example, the first principal eigenface), we will slide the template across an image with faces in it. We are interested in determining the potential matches between the template face and parts of the image in order to locate faces in the image. Which mathematical operation would be appropriate to use as a metric of similarity here?

Correct Answer

dot product

element-wise multiplication

element-wise addition

outer product

Multi-Class Voting

In lecture, we looked at the case of boosting for binary decisions. One of the underlying ways in which we combined weak learners was through the concept of majority voting. Here, we are interested in seeing if voting--in the way we previously defined it--can generalize to the case of tri-nary classification (i.e., classifying a sample into one of three classes). We will define a weak classifier as LaTeX: h\_i(x) \in \{-1, 0, +1\}h i ( x ) ∈ { − 1 , 0 , + 1 }, where -1, 0, +1 represent each of our three classes. The ensemble classifier will then be defined as LaTeX: H(x) = \textrm{sign}{(\Sigma\_{i} h\_i(x))}H ( x ) = sign ( Σ i h i ( x ) ). Will this voting method provide valid results when used in the context of boosting?

True

Correct Answer

False

Second Question Group Pick 5 questions, 1 pts per question

Integral Images

If LaTeX: AA is an image of size LaTeX: N\times MN × M pixels, with pixel values ranging from zero to one. Which of the following statements is True?

The complexity of computing its integral image is LaTeX: O(NM)O ( N M ).

The maximum value in its integral image will be LaTeX: \le NM≤ N M.

Integral images can be readily used to compute the eigenfaces in image LaTeX: AA.

Correct Answer

1, 2

only 1

only 2

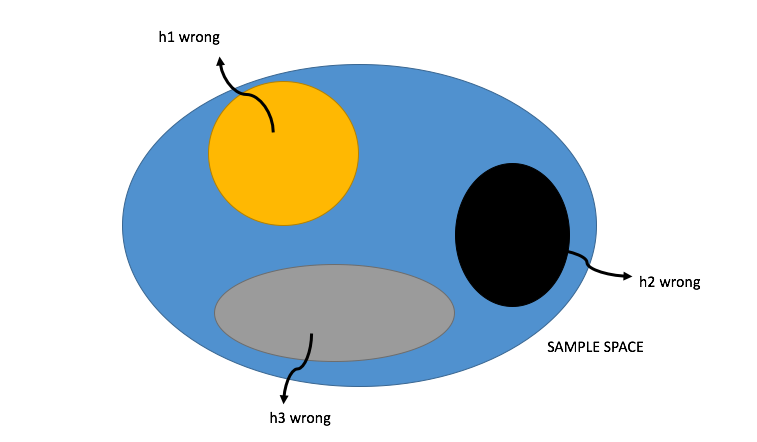
only 3

1, 2, 3

Subspaces and Errors

Suppose we have three weak learners, LaTeX: h\_1(x), h\_2(x), \textrm{and } h\_3(x)h 1 ( x ) , h 2 ( x ) , and h 3 ( x ) , the picture shown below represents the sample space and subspaces in which each classifier makes mistakes.

The sample space of an image with mutually exclusive regions demarcated. The three regions refer to the area of the sample space on which a given classifier is wrong. There is no overlap in the regions. Each classifier has its own region.



Which of the following statements are true:

1. We can always get the correct classification for any sample in the sample space using AdaBoost with these three weak learners.
2. LaTeX: h\_1(x), h\_2(x), \textrm{and } h\_3(x)h 1 ( x ) , h 2 ( x ) , and h 3 ( x ) are weak learners which means that their error rates are larger than 0.5.
3. LaTeX: H(x) = \textrm{sign}(h\_1(x)+h\_2(x)+h\_3(x))H ( x ) = sign ( h 1 ( x ) + h 2 ( x ) + h 3 ( x ) ) is a form of predicting the correct class via boosting

Correct Answer

1, 3

1, 2

2, 3

only 3

Boosting and Learners

Which of the following statements are true about boosting?

Correct Answer

Boosting can combine many weak classifiers to construct a strong classifier.

Correct Answer

Every weak learner focuses on failed classifications of the previous learner.

Weights of correctly classified samples are not changed during the training process.

Adding weak learners after the strong learner achieves zero errors may cause overfitting.

More Boosting and Learners

Which of following is true about boosting?

Correct Answer

Cross validation can be used to select the number of iterations in boosting; this procedure may help reduce overfitting.

We learn a classifier f by boosting weak learners h. The shape of f’s decision boundary is the same as h’s, but with different parameters. (e.g., if h was a linear classifier, then f is also a linear classifier).

All the samples are treated equally in learning the boosting classifier.

Once we get 100% accuracy on training set, we may not be able to improve the classifier anymore.

Cleaning Up Decisions

By performing the Viola-Jones algorithm on small patches of an image, it will often happen that a face is recognized but that there are several locations directly around the face that are similarly classified as faces. It is important to consider schemes that would combine these classifications into a single correct classification.

Correct Answer

True

False

Alternative Alphas

Which of the following functions could be viable alternatives for computing the classifier weights, LaTeX: \alpha\_tα t?

Correct Answer

LaTeX: \alpha\_t = \cot\left( \pi \varepsilon\_t \right)α t = cot ⁡ ( π ε t )

Correct Answer

LaTeX: \alpha\_t = \frac{1 - 2\varepsilon\_t}{\varepsilon\_t}α t = 1 − 2 ε t ε t

LaTeX: \alpha\_t = -\tan\left( \pi \varepsilon\_t - \frac{1}{2} \right)α t = − tan ⁡ ( π ε t − 1 2 )

LaTeX: \alpha\_t = \frac{-2\varepsilon\_t}{\varepsilon\_t-0.5}α t = − 2 ε t ε t − 0.5

Performance Curves 2

Near the end of lecture, we provided a modification to our standard boosted classifier, adding in a term to adjust classification: LaTeX: H\_{\textrm{modified}} = \textrm{sign}(\Sigma\_t \alpha\_t h\_t(x) + Y)H modified = sign ( Σ t α t h t ( x ) + Y ). Here, LaTeX: YY, is a real-valued parameter. If we continually decrease LaTeX: YY so that it is a large negative number, what do you expect to see happen to the performance of the classifier:

N.B.: There are two questions like this one. They test two different situations. Please read the question carefully.

Correct Answer

The false rejection rate will increase.

The false rejection rate will decrease.

The True Positive Rate (also known as sensitivity or recall) will increase.

Correct Answer

The True Positive Rate (also known as sensitivity or recall) will decrease.

The False Positive Rate will increase.

Correct Answer

The False Positive Rate will decrease.

Viola Jones 2

Which techniques/ideas reduced the number of required computations in the Viola-Jones algorithm? (Hint: it would help to read their paperLinks to an external site. )

Correct Answer

Attentional Cascade

Correct Answer

Integral Image

Rectangular features

AdaBoost

Weights and Errors of AdaBoost

Which of the following statements is true about Adaboost?

The total error, LaTeX: \varepsilon\_tε t, of a weak classifier is calculated as the ratio of the number of misclassified samples to the total number of samples.

The weights of the misclassified samples in a given iteration are scaled by the same factor.

The only real requirement for selecting a weak classifier is that it is right more than 50% of the time.

Correct Answer

Only 3

1, 2, and 3

1, 2

1, 3