CHUNYUAN LI

JIARUI XING

# CSE 517A Machine Learning THW2

### Problems

* + - 1. **(30 points) Naïve Bayes**

1. We have different s under 2 different categories s, so in total we need parameters for .
2. This time we need parameters.
3. If we make the naïve assumption, for each dimension we need only to estimate and , therefore in total we need parameters, which is much smaller than that of not using naïve assumption ().

Generally speaking, the naïve assumption can be seen as an approximation method which makes our problem easier to solve. So when the dimension is high, we’d better use it or the problem is almost intractable for having too much (exponential # of) parameters to estimate. However, in lower dimension where we are able to estimate all the parameters, we can get more precise result by not using it.

* + - 1. **(30 points) Naïve Bayes, Part II**

1. Use as feature vector of document, we have:

Therefore,

1. (i) Assuming we have in total samples, has different categories and has different classes,

(ii) The MLE estimators for Gaussian distribution are:

(iii) We have 2 classes, and for each we need to estimate means, variances and 1 prior , so there should be parameters.



As a result,

* + - 1. **(40 points) Valid Kernels, Kernel Construction**

1. Here . Define kernel function

We can see that transforming to needs 12 productions, and compute need additional 10 productions and 9 summations, thus in total we need 22 productions and 9 summations. However, compute need only 4 productions and 1 summation and can get the same result, which is highly efficient.

1. We can judge from the eigenvalues of matrices.
2. For , its eigen vectors and eigenvalues are:

Therefore, is positive semidefinite.

1. For , its eigenvectors and eigenvalues are:

Therefore, is positive definite.

1. For , its eigenvectors and eigenvalues are:

Therefore, is neither positive definite or semidefinite.

1. (i) Assuming A and B are both matrices, since A and B are both positive semi-definite matrices, we have:

Therefore,

Thus (A+B) is positive semi-definite.

(ii) Using property of Kronecker product [1]:

Suppose that A and B are square matrices of size n and m respectively. Let be the eigenvalues of A and be those of B (listed according to multiplicity). Then the eigenvalues of A ⊗ B are

Since A and B are both positive semi-definite, we know:

Therefore, , i.e. all eigenvalues of A ⊗ B nonnegative. As a result A ⊗ B is positive semi-definite.

1. By Taylor expansion:

Where

Which is a vector in infinite dimensional space.

References:

[1] [Kronecker product - Wikipedia](https://en.wikipedia.org/wiki/Kronecker_product#Abstract_properties)

* + - 1. **(30 points) Kernelize the Perceptron Algorithm**

1. Note where is the number of times was misclassified, then:
2. From (a):

|  |
| --- |
|  |

If we use feature transformation , and kernel function , then:

1. The kernelized version of perceptron is:
2. Initialize

REPEAT until convergence:

1. Pick randomly from
2. If , update :