5. Bayesian Information Criterion

So far we have simply set the number of mixture components K but this is also a parameter that we must estimate from data. How does the log-likelihood of the data vary as a function of K assuming we avoid locally optimal solutions?

To compensate, we need a selection criterion that penalizes the number of parameters used in the model. The Bayesian information criterion (BIC) is a criterion for model selection. It captures the tradeoff between the log-likelihood of the data, and the number of parameters that the model uses. The BIC of a model M is defined as:

$$\text{BIC}\left(M\right) = l - \frac{1}{2}p\log n$$

where l is the log-likelihood of the data under the current model (highest log-likelihood we can achieve by adjusting the parameters in the model), p is the number of free parameters, and n is the number of data points. This score rewards a larger log-likelihood, but penalizes the number of parameters used to train the model. In a situation where we wish to select models, we want a model with the the highest BIC.

Implementing the Bayesian Information Criterion

0.0/1.0 point (graded)

Fill in the missing Bayesian Information Criterion (BIC) calculation (`bic` function) in `common.py` .

Available Functions: You have access to the NumPy python library as `np` , to the `GaussianMixture` class and to typing annotation `typing.Tuple` as `Tuple` .

```
def bic(X: np.ndarray, mixture: GaussianMixture,
        log_likelihood: float) -> float:
    """Computes the Bayesian Information Criterion for a
    mixture of gaussians

    Args:
        X: (n, d) array holding the data
        mixture: a mixture of spherical gaussian
        log_likelihood: the log-likelihood of the data

    Returns:
        float: the BIC for this mixture
    """
    raise NotImplementedError

```

Press ESC then TAB or click outside of the code editor to exit

Unanswered

Generating Speech Output

```
def bic(X: np.ndarray, mixture: GaussianMixture,
        log_likelihood: float) -> float:
    """Computes the Bayesian Information Criterion for a
    mixture of gaussians

    Args:
        X: (n, d) array holding the data
        mixture: a mixture of spherical gaussian
        log_likelihood: the log-likelihood of the data

    Returns:
        float: the BIC for this mixture
    """
    n, _ = X.shape
    K, d = mixture.mu.shape

    return log_likelihood - (K * d + K + K - 1) / 2 * np.log(n)
```

Solution:

The Bayesian Information Criterion for a mixture of spherical Gaussians is:

$$BIC\left(D;\theta\right) = l\left(D;\theta\right) - \frac{k\left(d+2\right)-1}{2}\log\left(n\right)$$

Submit You have used 0 of 25 attempts

Answers are displayed within the problem

Picking the best K

0.0/1.0 point (graded)

Find the best K from $[1, 2, 3, 4]$ on the toy dataset. This will be the K that produces the optimal BIC score. Report the best K and the corresponding BIC score. Measure the BIC on EM models, only. Does the criterion select the correct number of clusters for the toy data?

Generating Speech Output

Best K = **Answer:** 3

Best BIC =

Answer: -1169.2589

Grader note: While the best BIC should be a negative value, due to earlier grader error, we have corrected the grader to accept both the positive and the negative value.

Solution:

Code:

```
def run_with_bic():
    max_bic = None
    for K in range(1, 5):
        max_ll = None
        best_seed = None
        for seed in range(0, 5):
            mixture, post = common.init(X, K, seed)
            mixture, post, ll = naive_em.run(X, mixture, post)
            if max_ll is None or ll > max_ll:
                max_ll = ll
                best_seed = seed

        mixture, post = common.init(X, K, best_seed)
        mixture, post, ll = naive_em.run(X, mixture, post)
        bic = common.bic(X, mixture, ll)
        if max_bic is None or bic > max_bic:
            max_bic = bic
        title = "EM for K=, seed=, ll=, bic=".format(K, best_seed, ll,
bic)
        print(title)
        common.plot(X, mixture, post, title)
```

Generating Speech Output

1	-1315.5056
2	-1195.0397
3	-1169.2589
4	-1180.0121

From the BIC values above, the best K is 3, which seems to fit the toy data.

Submit You have used 0 of 10 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Project 4: Collaborative Filtering via Gaussian Mixtures / 5. Bayesian Information Criterion

Add a Post

Show all posts by recent activity

My unit tests for checking BIC function
I've updated my tests from part 3 to include two tests for BIC. See [here][1] for tests. There is ... 4
Pinned Community TA

What is the purpose of this exercise
This was NOT covered in the lecture. Neither does it come close to the theory covered in the ... 2

Decimal off in my output, any hint?
Used the formula and the right parameters (ll,p,size) but I am still off My Output: bix:-1218.**... 5

Projects absolutely beyond the lectures....
Hi MIT staff, I'm quite frustrated since I've found lectures far, far away from projects demand... 6

Generating Speech Output

?	BIC vs AIC When would we use the Bayesian Information Criterion instead of the Aikake Information Cri...	2
☑	BIC Definition??? Good day! According to BIC definition, I use "number of adjustable parameters", counting the...	9
☑	Can Someone explain free variabele? I managed to solve the question, but have no clue what I am doing. Any source/link would be...	5
☑	Picking the best K marked as wrong although all previous exercises marked as correct Hello there, I got the BIC implementation right. Also all the previous exercises were marked a...	2
💬	Replacing "adjustable" with "free" avoids a lot of confusion Community TA	5
?	BIC I used the X values from the incorrect output my code produced, but when I run it locally wit...	4
💬	[Staff] Could not format HTML for problem. Contact course staff in the discussion forum for assistance. Message appears at the bottom of the page.	4