EM Question Pick 10 questions, 1 pts per question

Your friendly neighborhood gamblers

In class, we gave a dice shooter example to illustrate Expectation Maximization(EM) algorithm (slide 129), about this example, which of the following statements are true?

In the initialization step, it's necessary to initialize LaTeX: P(k=red|o)P ( k = r e d | o ), LaTeX: P(k=blue|o)P ( k = b l u e | o ), and LaTeX: P(o)P ( o )

Iterations will keep going until LaTeX: P(o)P ( o )converges

We will find unique LaTeX: P(o)P ( o ) when it converges

Correct Answer

only 2

1 and 3

1 and 2

2 and 3

Gradient Descent

In class, we showed that the EM algorithm can be used to maximize the likelihood function. Why did we not choose gradient descent instead?

EM and gradient descent are equivalent and both end up with the same solution once they converge

Gradient descent does not guarantee to maximize the likelihood function at each iteration

In some cases where the likelihood function has integrals, computing the gradients is intractable

Correct Answer

only 2

2 and 3

1 and 2

1, 2, and 3

Convergence of EM

True or False: The EM algorithm ensures to never decrease the objective function at any iteration.

Correct Answer

True

False

Number of parameters

Suppose our observations are composed of 40-dimensional vectors and we assume that these observations have been generated by a model composed of a mixture of 13 Gaussians with diagonal covariance. If we were to use EM to learn the model, how many parameters we would have to learn?

Correct Answer

1053

1040

1055

1060

Conditions for q

The variational lower bound is defined as:

LaTeX: log \sum\_{h}{Q(h)\frac{P(h,o)}{Q(h)}} \ge \sum\_{h}{Q(h) log \frac{P(h,o)}{Q(h)}}l o g ∑ h Q ( h ) P ( h , o ) Q ( h ) ≥ ∑ h Q ( h ) l o g P ( h , o ) Q ( h )

and the choice of LaTeX: Q(h)Q ( h ) can be arbitrary as long as LaTeX: Q(h)Q ( h ) satisfies which of the following conditions?

LaTeX: Q(h)Q ( h ) is concave

LaTeX: Q(h)\ge 0 Q ( h ) ≥ 0 for all LaTeX: hh

LaTeX: \sum\_{h}{Q(h)} = 1∑ h Q ( h ) = 1

Correct Answer

2 and 3

1 and 3

1 and 2

1, 2, and 3

Probability Function

If LaTeX: p\_1p 1 and LaTeX: p\_2p 2 are two probability functions and LaTeX: \lambda\_1, \, \lambda\_2λ 1 , λ 2 real numbers, then LaTeX: \lambda\_1 p\_1 + \lambda\_2 p\_2λ 1 p 1 + λ 2 p 2 is a probability function if:

LaTeX: \lambda\_1λ 1 and LaTeX: \lambda\_2λ 2 are positive

LaTeX: \lambda\_1 + \lambda\_2 = 1λ 1 + λ 2 = 1

LaTeX: \lambda\_1 \cdot \lambda\_2 = 1λ 1 ⋅ λ 2 = 1

Correct Answer

1 and 2

only 1

only 2

1 and 3

ELBO

True of False:

In order to ensure that maximizing the variational lower bound LaTeX: J(\theta,\theta')J ( θ , θ ′ ) will maximize the likelihood LaTeX: log P(o;\theta)l o g P ( o ; θ ), we require LaTeX: log P(o;\theta) = J(\theta, \theta)l o g P ( o ; θ ) = J ( θ , θ )

Correct Answer

True

False

EM algorithm

We prefer to use the EM algorithm because:

It is much faster than other methods

Maximizing directly the likelihood is an intractable problem

EM converges to a unique solution

EM guarantees non-decreasing likelihood across iterations

only 2

only 3

1 and 2

2 and 3

Correct Answer

2 and 4

EM Properties

Which of the following statements are true about Expectation Maximization?

The expectation step is followed by the maximization step

In the initialization step, we can either randomly assign starting values from a constrained distribution or empirically compute them from the data

The efficiency of the EM estimation is affected by the number of latent classes

Correct Answer

1, 2 and 3

1

1 and 2

3

MLE

Which of the following is FALSE?

It is not always possible to have a closed-form solution of the maximum likelihood estimate

In our maximum likelihood definition of "best-fit", we assume the observed data is not representative of the underlying process

In maximum likelihood estimate, we aim to find the parameters of a distribution that maximizes the probability of generating the observations

Dirichlet distribution has a closed-form maximum likelihood estimate

Correct Answer

2 and 4

1 and 2

2 and 3

None of them

Incomplete data

Which of the following is CORRECT?

When we use MLE to estimate the missing components, we marginalize over missing components and observations

In EM, we fill the missing components by sampling from the posterior probability distribution

When EM is used to fill in missing components, we don't require the model parameters to be initialized

Correct Answer

2

3 and 1

2 and 3

3