Parameter Selection Process WF

Thursday, July 25, 2024 3:20 PM

Parameter Selection Process Notes and Sources:
Richie Fleming

Methods in Order:

- Bayesian Parameter Selection Method

Bayesian Selection Method:

Terms I need to Learn for method:

- Markov Chain Monte Carlo Sampling Method

Markov Chains:

(also watch hidden Markov Video)

Ex.)

- Every day we can have one of two states (Sunny or Cloudy)
- Want to find prob of sunny the next day
- This approach is governed by 4 different outcomes (S to S, S to C, C to S and C to C)

Markov Assumption:

- The weather on any given day only depends on the weather one day prior. P(Wt | Wt-1...Wt-n) is the same as P(Wt | Wt-1)
- This can be made into a matrix.

Steady State Markov Chains:

Ex)

- D0 100% its sunny, cloudy = 0 %
- What the prob on the next day.
- Keep iterating using .3 and .7 know probs from pervious example
- Do we ever converge to a steady state?
- Assume that there is a convergence, steady state moving forward.
- Steady states of sunny and cloudy need to add up to 1.
- Assuming steady states we can calculate Pi1 * .3 + Pi2 * 0.5 =1 and solve for Pi1 and Pi2 -> averages prob of sunny and cloudy days.

N X N matrix

- Can use this in NLP as well ** (use for prompts)
- What's the highest prob move from Baltic Ave (Monopoly)

Monte Carlo Method:

Ex1.) Throwing darts into a square

- Randomly thrown, at square with smaller circle/dartboard in square
- What is the prob we hit the circle when thrown
- Area of circle / area of the squared = Prob (pi/4)
- Different method: using simulation strategy.
- Start by simulating N dart throws

- Area of circle / area of the squared = Prob (pi/4)
- Different method: using simulation strategy.
- Start by simulating N dart throws
- How many points live in the circle / N is the prob using sim method.
- More samples we run, the closer we get to actually distribution.

Ex2.)

-

MCMC:

- Sample from P(x) -> although we might not know the denominator, so we only have the numerator, f(x)
- Make sure our g(x) function lies above f(x) (we know g(x))
- ^^^^^^^^^^^^^^This is for accept-reject^^^^^^^^^^^^

Issue with accept reject:

- It is less applicable to distributions in the real world.
- Inefficient sampling
- Samples are uncorrelated

Core:

- Learn from previous sample, pick the next sample based on the information gathered (Markov) -> Next sample depends on previous.
- Monte Carlo: Simulating Markov Chain
- The idea is to simulate draws from P(x) ->PDF

How Do we Design Markov Chain?

How do we define some rule to get the next sample?

- Find Steady State of Markov Chain
- Sampling from the steady state distribution
- After we use the first couple values to determine the steady state, we don’t need them in the future (more computational power in the first calculation leads to less in future calculations) ->"Burn" those values

Design Transition Probs:

- p(x) T(y |x) = p(y) T(x|y) -> Detailed balance condition, if true then this p(x) is the stationary state of the Markov chain.
- For any states (x,y)
- T(y|x), conditional prob of going to state y given previous state of x
- pT = P in matrix form, where T is transition matrix

Gibbs Sampling:

Goal: Sample for 2D normal distribution

- Gibbs sampling is useful when we have 2 or more dimensions

Use Gibbs When:

- Joint PDF: sampling is difficult
- Sampling from conditional distr is easy (p(x|y) or p(y|x))

Algo:

1.) Initialize at x0, y0

2.) Sample x1 ~ p(x1 | y1)

3.) Sample y1 ~ p(y1 | x1)

Explanation: Get new sample from changing both x and y

Issues:

- Probability Spikes: Low -> Low, High -> High
- High prob surrounded by low prob, or vice versa