Linear Regression Analysis
- Bayesian Approach -

Tseng-Ching James Shen, PhD

Regression Model

- The parameters $w_0, w_1, w_2, \dots w_d$ are random variables

$$y = w_0 + w_1 x_1 + \dots + w_d x_d$$

- We will need to estimate the joint distribution of these parameters given by the observed dataset point, *Posterior*

$$p(\mathbf{w}|\mathbf{X},\mathbf{y})$$

- It would be difficult to directly compute the posterior
- Bayesian approach is commonly used in machine learning to estimate the posterior

Bayesian Approach

- We can apply the Bayer's rule to re-write the distribution as

  $p(\mathbf{w}|\mathbf{X},\mathbf{y})$ =

  where

  =

- We will need to use the following to calculate the posterior
  - Likelihood
  - Prior
- There are two ways for the calculation
  - Conjugate prior-likelihood
  - Sampling Technique

Conjugate Prior-Likelihood

- If the prior and likelihood distributions are conjugate, the posterior will be of the same form as the prior
- Scenario 1

  Prior – Beta distribution

  Likelihood – Binomial distribution

  => Posterior – Beta distribution
- Scenario 2 (*)

  Prior – Gaussian distribution

  Likelihood – Gaussian distribution

  => Posterior – Gaussian distribution

Scenario 2 – Gaussian Conjugate

- Assumptions
  - Likelihood

    $N(\mathbf{X}\,\mathbf{w}, \;)$
  - Prior

    $N(\boldsymbol{\mu}_0, \;)$
- The posterior can be derived as (sec 3.8 in textbook)

  $p(\mathbf{w}|\mathbf{X},\mathbf{y}) = N(\boldsymbol{\mu}_w, \;)$

  where

  $= (\mathbf{X}^T\mathbf{X} + {}^{-1})^{-1}$

  $\mu_w = (\mathbf{X}^T\mathbf{y} + {}^{-1}\,\boldsymbol{\mu}_0)$

Sampling

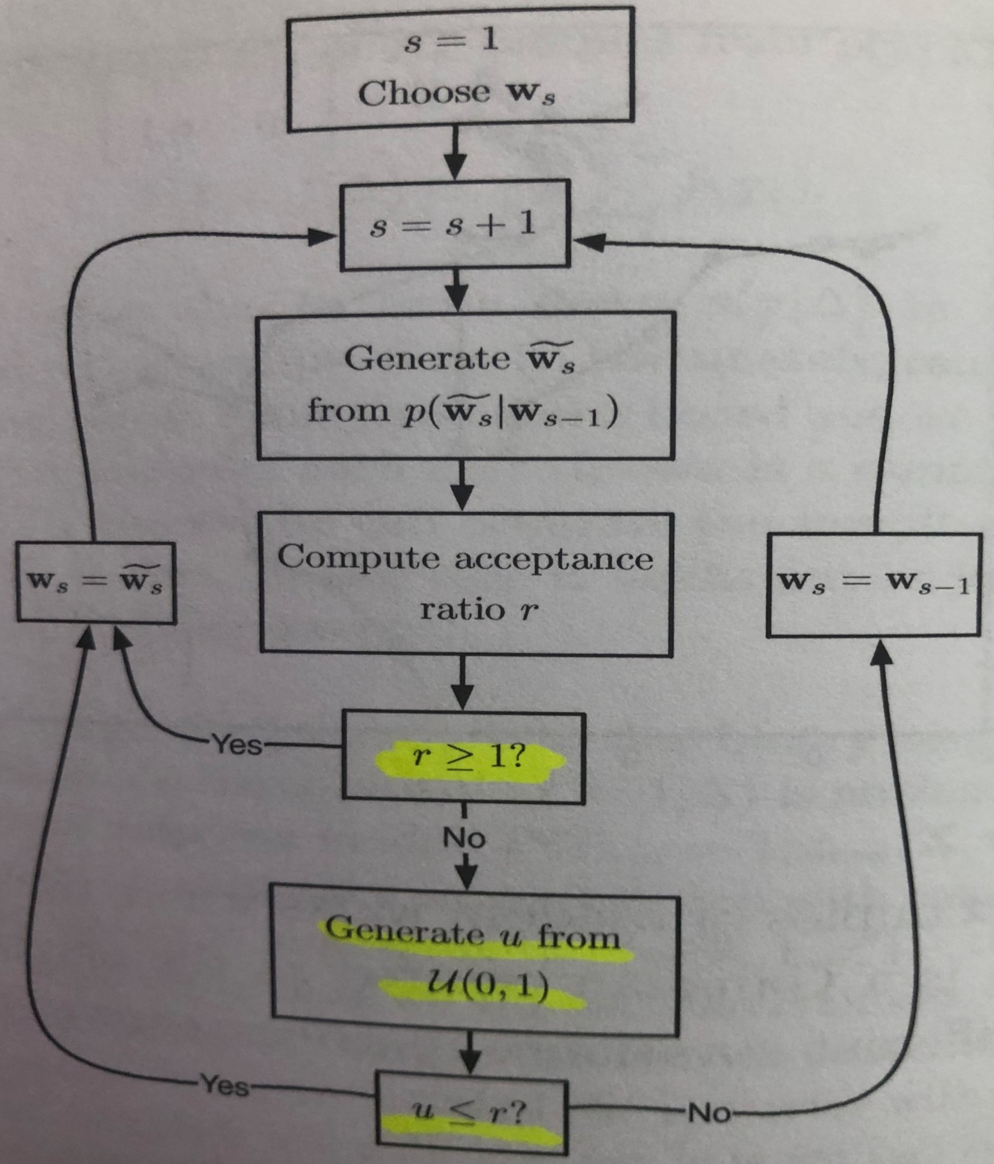
- Sampling from a know distribution
- Sampling from an unknown distribution
- From observed dataset points, use a sampling technique e.g. the *Metropolis-Hastings algorithm* to derive samples of the unknown distribution
- The algorithm implements an iterative process which produces a sequence of samples $\mathbf{w}^{(1)}$, $\mathbf{w}^{(2)}$, …, $\mathbf{w}^{(s)}$
- At i-th iteration of the iterative process, the technique does the following:
  - ➢Propose a new sample with a proposal density function, $N(\mathbf{w}^{(i-1)},)$
  - ➢Use a *criteria* to decide if the proposed sample should be accepted or rejected and determine the new sample using the following rules
    - ○ If accepted, $\mathbf{w}^{(i)} =$
    - ○ If rejected, $\mathbf{w}^{(i)} = \mathbf{w}^{(i-1)}$

Accepting Criteria

- Assumption: $p(\mathbf{w})$ = N($\mathbf{0}$,)
- With , calculate the ratio as follows

  r =

- The accepting/rejecting condition
  - If r
  - If r < 1, then generate a value u from a uniform distribution between 0 and 1,
    - If u then accept the proposed sample
    - If u > r, then reject the proposed sample



The Metropolis–Hastings algorithm

PyCM3

- Probabilistic programming framework for Python
- Provides API to specify the Bayesian model
- Implements various Markov Chain Monte Carlo (MCMC) sampling algorithms including Metropolis-Hastings algorithm
- Can be used to implement the Bayesian regression analysis