Student Name: Guangyu Lin

Collaboration Statement:

Total hours spent: 4 hours

I discussed ideas with these individuals:

- TODO
- TODO
- ...

I consulted the following resources:

- office hour with Professor
- TODO
- ...

By submitting this assignment, I affirm this is my own original work that abides by the course collaboration policy.

Links: [HW1 instructions] [collab. policy]

Contents

1a: Problem Statement

Let $\rho \in (0.0, 1.0)$ be a Beta-distributed random variable: $p \sim \text{Beta}(a, b)$.

Show that $\mathbb{E}[\rho] = \frac{a}{a+b}$.

Hint: You can use these identities, which hold for all $a > 0$ and $b > 0$:

$$\Gamma(a) = \int_{t=0}^{\infty} e^{-t} t^{a-1} dt \tag{1}$$

$$\Gamma(a+1) = a\Gamma(a) \tag{2}$$

$$\int_0^1 \rho^{a-1}(1-\rho)^{b-1} d\rho = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{3}$$

1a: Solution

$\mathbb{E}[\rho]$

= $\mathbb{E}[Beta(a, b)]$ substitute the beta pdf by definition

= $\Sigma \mu c(a, b) \mu^{a-1}(1-\mu)^{b-1}$ move the $c(a, b)$ outside of the function

=$c(a, b)\Sigma \mu^a \mu^{b-1}$ by gamma function's identity 3

= $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+1) \cdot \Gamma(b)}{\Gamma(a+b+1)}$ by using the second identity

= $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)}$ by cancel some terms we can get

=$\frac{a}{a+b}$

1b: Problem Statement

Let μ be a Dirichlet-distributed random variable: $\mu \sim \text{Dir}(a_1, \ldots a_V)$.

Show that $\mathbb{E}[\mu_w] = \frac{a_w}{\sum_{v=1}^{V} a_v}$, for any integer w that indexes a vocabulary word.

** Hint:** You can use the identity:

$$\int \mu_1^{a_1-1}\mu_2^{a_2-1}\dots\mu_V^{a_V-1}d\mu = \frac{\prod_{v=1}^{V}\Gamma(a_v)}{\Gamma(a_1+a_2\dots+a_V)} \tag{4}$$

1b: Solution

$\mathbb{E}[\mu_w]$

= $\Sigma\mu$DirPDF($\mu|a$) substitute Dirichlet distribution by its definition

= $\Sigma c(a)\prod_{v=1}^{V}\mu^{a_v}$ move the c(a) outside of the function because it is a constant

= $c(a)\Sigma\prod_{v=1}^{V}\mu^{a_v}$ using Dirichlet distribution's identity

= $\frac{\Gamma(\Sigma a_v)}{\prod_v\Gamma(a_v)}\cdot\frac{\prod\Gamma(a_v+1)}{\Gamma(\Sigma a_v+1)}$ using gamma function identity

=$\frac{\Gamma(\Sigma a_v)}{\prod_v\Gamma(a_v)}\cdot\frac{a_v\prod\Gamma(a_v)}{\Sigma a_v\Gamma(\Sigma a_v)}$ cancel some terms and we get

=$\frac{a_v}{\Sigma a_v}$

2a: Problem Statement

Show that the likelihood of all N observed words can be written as:

$$p(X_1=x_1,X_2=x_2,\dots,X_N=x_N|\mu)=\prod_{v=1}^{V}\mu_v^{n_v} \tag{5}$$

2a: Solution

$p(X_1=x_1,...,X_n=x_n|\mu)$

= $\prod_{n=1}^{N}CatPMF(X_n|\mu)$ by the definition of categorical distribution

= $\prod_{n=1}^{N}\prod_{v=1}^{V}\mu^{X_{nv}}$ because $n_v=\Sigma_{n=1}^{N}[X_n=V]$ so we can get

=$\prod_{v=1}^{V} \mu_v^{n_v}$

2b: Problem Statement

Derive the next-word posterior predictive, after integrating away parameter μ.

That is, show that after seeing the N training words, the probability of the next word X_* being vocabulary word v is:

$$p(X_* = v | X_1 = x_1 \ldots X_N = x_n) = \int p(X_* = v, \mu | X_1 = x_1 \ldots X_N = x_n) d\mu$$
$$= \frac{n_v + \alpha}{N + V\alpha} \quad (6)$$

2b: Solution

$\int p(X_* = v, \mu | X_1 = x_1 \ldots X_N = x_n) d\mu$ using product rule we can decompose it into two part

= $\int p(X_* | \mu, X_1, \ldots, X_n) p(\mu | X_1, \ldots, X_n) d\mu$ the first part can be written as $p(x_* | \mu)$ and the second part can be derived as DirPDF due to the conditional independence

=$\int Cat(x_* = v | \mu) \cdot DirPDF(\mu | \alpha + n) d\mu$ substitute by each distribution's definition

=$\int \mu_v \cdot c(\alpha + n) \cdot \prod_{v=1}^{V} \mu_v^{\alpha_v + n_v - 1} d\mu$ moved the constant part outside of the integral

=$c(\alpha + n) \int \mu_v \cdot \prod_{v=1}^{V} \mu^{\alpha_v + n_v - 1} d\mu$ we can create β as a new vector that contains $[\alpha_1, \alpha_2, ..., \alpha_v + 1, \alpha_V]$ then we have

=$c(\alpha + n) \int \mu_v \cdot \prod_{v=1}^{V} \mu^{(\beta_v + n_v - 1)} d\mu$ then we can use the identity of dirichlet distribution

= $\frac{\Gamma(\Sigma_v \alpha_v + n_v)}{\prod_v \Gamma(\alpha_v + n_v)} \cdot \frac{\prod_v \Gamma(\beta_v + n_v)}{\Gamma(\Sigma_v \beta_v + n_v)}$ then we substitue the β_v as $\alpha_v + 1$ and using the identity of gamma function we can get

= $\frac{\Gamma(\Sigma_v \alpha_v + n_v)}{\prod_v \Gamma(\alpha_v + n_v)} \cdot \frac{(n_v + \alpha_v) \prod_v \Gamma(\alpha_v + n_v)}{\Sigma_v \alpha_v + n_v \Gamma(\Sigma_v \alpha_v + n_v)}$ we can cancel some terms and $\Sigma_v \alpha_v$ can be written as $V\alpha$ and $\Sigma_v n$ can be written as N and we get

= $\frac{n_v + \alpha}{N + V\alpha}$

2c: Problem Statement

Derive the marginal likelihood of observed training data, after integrating away the parameter μ.

That is, show that the marginal probability of the observed N training words has the following closed-form expression:

$$p(X_1 = x_1 \ldots X_N = x_N) = \int p(X_1 = x_1, \ldots X_N = x_N, \mu) d\mu \tag{7}$$

$$= \frac{\Gamma(V\alpha) \prod_{v=1}^{V} \Gamma(n_v + \alpha)}{\Gamma(N + V\alpha) \prod_{v=1}^{V} \Gamma(\alpha)} \tag{8}$$

2c: Solution

$\int p(X_1 = x_1, ..., X_n = x_n, \mu) d\mu$ using Bayes Rule we can get

= $\int p(X_1 = x_1, ..., X_n = x_n | \mu) \cdot p(\mu) d\mu$ substitute the first term by the result of 2a and the second part is just the dirichelet distribution

=$\int \prod_{v=1}^{V} \mu_v^{n_v} \cdot DirPDF(\mu|\alpha) d\mu$ so we can substitue with the definition of dirichlet distribution and we can get

= $\int \prod_{v=1}^{V} \mu_v^{n_v} \cdot c(\alpha) \cdot \prod_{v=1}^{V} \mu_v^{\alpha_v - 1}$ we can move the constant part outside and combine these two terms

=$c(\alpha) \int \prod_{v=1}^{V} \mu_v^{\alpha_v + n_v - 1} d\mu$ then we can use the identity of dirichlet distribution and get

= $\frac{\Gamma(\Sigma_v \alpha_v)}{\prod_{v=1}^{V} \Gamma(\alpha_v)} \cdot \frac{\prod_{v=1}^{V} \Gamma(\alpha_v + n_v)}{\Gamma(\Sigma_v \alpha_v + n_v)}$ then $\Sigma_v \alpha_v$ can be written as $V\alpha$ and $\Sigma_v \alpha_v + n_v$ can be written as $V\alpha + N$ then we get

$$= \frac{\Gamma(V\alpha)\cdot\prod_{v=1}^{V}\Gamma(\alpha_v+n_v)}{\Gamma(N+V\alpha)\cdot\prod_{v=1}^{V}\Gamma(\alpha_v)}$$