Assignment 1 for #70240413
"Statistical Machine Learning"

Kui XU, 2016311209

2017/03/23

1 Mathematics Basics

Choose one problem from the 1.1 and 1.2. A bonus would be given if you finished the both.

1.1 Calculus

The gamma function is defined by (assuming x > 0)

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du. \tag{1}$$

(1) Prove that $\Gamma(x+1) = x\Gamma(x)$.

(2) Also show that

$$\int_0^1 u^{a-1}(1-u)^{b-1} du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{2}$$

Solution: For Question (1), we can prove it by Using integration by parts, the steps are as follows:

$$\begin{aligned}
\Gamma(x+1) &= \int_0^\infty u^x e^{-u}\, du \\
&= \left[-u^x e^{-u}\right]_0^\infty + \int_0^\infty x u^{x-1} e^{-u}\, du \\
&= \lim_{u\to\infty}(-u^x e^{-u}) - (0e^{-0}) + x\int_0^\infty u^{x-1} e^{-u}\, du \\
&= x\int_0^\infty u^{x-1} e^{-u}\, du \\
&= x\Gamma(x)
\end{aligned} \tag{3}$$

As we know, when $u \to \infty$, $-u^x e^{-u} \to 0$, so the equation is proved.

Solution: For Question (2), we know that the left of the equation is a Beta function. From the definitions, we can express the equation which we want to prove as :

$$\Gamma(a+b)B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{4}$$

It's a double integral, the expansion formula is as follows:

$$\begin{aligned}
\Gamma(a+b)B(a,b) &= \int_0^\infty u^{a+b-1} e^{-u} du \int_0^1 v^{a-1}(1-v)^{b-1} dv \\
&= \int_0^\infty \int_0^1 (uv)^{a-1}[u(1-v)]^{b-1} u e^{-u} du\, dv
\end{aligned} \tag{5}$$

Then we do a transformation $w = uv$, $z = u(1-v)$. The inverse transformation is $u = w+z$, $v = w/(w+z)$, the corresponding ranges of them are $w \in (0, \infty)$ and $u \in (0, \infty)$.
The absolute value of the Jacobian is

$$\left| \nabla \frac{\partial(u,v)}{\partial(w,z)} \right| = \frac{1}{(w+z)} \tag{6}$$

Next, we use the changed of variables to do a double integral, the equation above becomes:

$$\begin{aligned}
&\int_0^\infty \int_0^\infty w^{a-1} z^{b-1} (w+z) e^{-(w+z)} \frac{1}{w+z} dw\, dz \\
&= \int_0^\infty \int_0^\infty w^{a-1} z^{b-1} e^{-(w+z)} dw\, dz \\
&= \int_0^\infty w^{a-1} e^{-w} dw \int_0^\infty z^{b-1} e^{-z} dz \\
&= \Gamma(a)\Gamma(b)
\end{aligned} \tag{7}$$

Finally the equation is proved.

1.2 Optimization

Use the Lagrange multiplier method to solve the following problem:

$$\begin{aligned}
\min_{x_1, x_2} \quad & x_1^2 + x_2^2 - 1 \\
s.t. \quad & x_1 + x_2 - 1 = 0 \\
& 2x_1 - x_2 \geq 0
\end{aligned} \tag{8}$$

Solution: Consider the above equation is consist of inequality constraint functions and it is a nonlinear optimization problem, we can use the lagrange multiplier method with KKT condition to solve it. The Objective funtion is

$$\mathcal{L}(x_1, x_2, \lambda, \mu) = x_1^2 + x_2^2 - 1 + \lambda \cdot \Big(x_1 + x_2 - 1\Big) + \mu \cdot \Big(2x_1 - x_2\Big) \tag{9}$$

The certain conditions which are called KKT condition should satisfy,

$$\begin{aligned}
\frac{\partial(L)}{\partial(X)}|_X &= 0 \\
\lambda_j &\neq 0 \\
\mu_k &\geq 0 \\
\mu_k \cdot \Big(x_1^* + x_2^* - 1\Big) &= 0 \\
x_1^* + x_2^* - 1 &= 0 \\
2x_1^* - x_2^* &\leq 0
\end{aligned} \tag{10}$$

Choose one problem from the following 1.3 and 1.4. A bonus would be given if you finished the both.

1.3 Stochastic Process

We toss a fair coin for a number of times and use H(head) and T(tail) to denote the two sides of the coin. Please compute the expected number of tosses we need to observe a first time occurrence of the following consecutive pattern

$$H, \underbrace{T, T, ..., T}_{k}. \tag{11}$$

Solution: we asume that E is the expection of the consecutive pattern $H,\underbrace{T,T,...,T}_{k}$, and E_T^k is the expection of $\underbrace{T,T,...,T}_{k}$. Consider an equivalent form of this pattern $H,\underbrace{T,T,...,T}_{k-1},T$, we have

$$\begin{cases} E = 1 + \frac{1}{2}E + \frac{1}{2}E_T^k, \\ E_T^k = E_T^{k-1} + 1 + \frac{1}{2}E_T^k + \frac{1}{2} \times 0. \quad E_T^1 = 2 \end{cases} \tag{12}$$

which $E = 1 + \frac{1}{2}E + \frac{1}{2}E_T^k$ shows the expection of the first toss. At the first time, you may get H or T with the $\frac{1}{2}$ probability. If you got H, OK, you succuced and then you will try to get k times T, the expection will be $\frac{1}{2}E_T^k$; If you got T, you fail and will restart to tosses and the expection will be $\frac{1}{2}E$. which $E_T^k = E_T^{k-1} + 1 + \frac{1}{2}E_T^k + \frac{1}{2} \times 0$ shows the the expection of the $k-1$ times of T (E_T^{k-1}) and the last toss. At the last toss, as for the first time, you will get H or T with the $\frac{1}{2}$ probability. If you got H, you fail and you need to get k times T over again and the expection will be $\frac{1}{2}E_T^k$. If you got T, OK, you win the game, the expection will be $\frac{1}{2}E_T^k$;
Next, we solve the recursive function above

$$E_T^k = 2^{k+1} - 2 \tag{13}$$

$\Rightarrow$

$$E = 1 + \frac{1}{2}E + \frac{1}{2}(2^{k+1} - 2) \tag{14}$$

$\Rightarrow$

$$E = 2^{k+1} \tag{15}$$

So the expected number of tosses is 2^{k+1}.

1.4 Probability

Suppose $p \sim Beta(p|\alpha, \beta)$ and $x|p \sim Bernoulli(x|p)$. Show that $p|x \sim Beta(p|\alpha + x, \beta + 1 - x)$, which implies that the Beta distribution can serve as a conjugate prior to the Bernoulli distribution.

Solution: Consider calculating the posterior $p|x$, and we know the likelihood function $x|p$ and the prior p, here we use Bayes' theorem:

$$\begin{aligned} P(p|x) &= \frac{P(x|p)P(p)}{P(x)} \\ &= \frac{P(x|p)P(p)}{\int P(x|p')P(p')dp'} \end{aligned} \tag{16}$$

From the definition, $P(p) \sim Beta(p|\alpha, \beta)$ and $P(x|p) \sim Bernoulli(x|p)$, and the Beta function is

$$Beta(p|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{17}$$

so $P(p|x)$ should be

$$
\begin{aligned}
P(p|x) &= \frac{P(x|p)P(p)}{\int_0^1 P(x|p')P(p')dp'} \\
&= \frac{\binom{m}{n} p^m (1-p)^{n-m} \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}}{\int_0^1 \binom{m}{n} p^m (1-p)^{n-m} \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1} dp} \\
&= \frac{p^{\alpha+m-1}(1-p)^{\beta-1+n-m}}{\int_0^1 p^{\alpha+m-1}(1-p)^{\beta-1+n-m} dp} \\
&= \frac{p^{\alpha+m-1}(1-p)^{\beta-1+n-m}}{B(\alpha+m, \beta+n-m)} \\
&= Beta(p|\alpha+m, \beta+n-m)
\end{aligned} \tag{18}
$$

So, it implies that the Beta distribution can serve as a conjugate prior to the Bernoulli distribution.

Differentiation allows for the calculation of the slope of the tangent of a curve at any given point as shown in Figure 1.

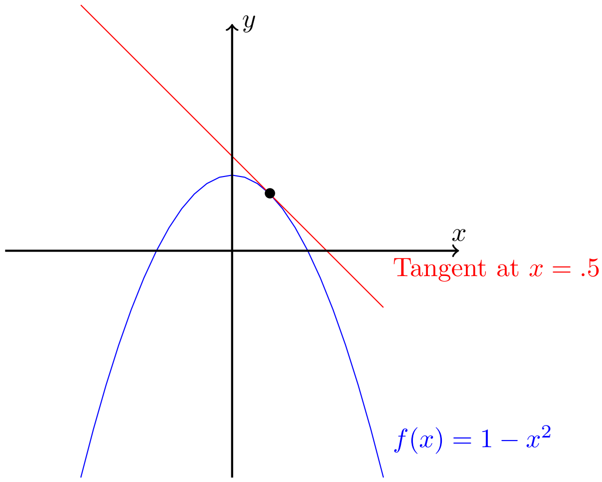

Figure 1: The plot of $f(x) = 1 - x^2$ with a tangent at $x = .5$.

Differentiation is now a technique taught to mathematics students throughout the world. In this document I will discuss some aspects of differentiation.

2 Exploring the derivative using Sage

The definition of the limit of $f(x)$ at $x = a$ denoted as $f'(a)$ is:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{19}$$

The following code can be used in sage to give the above limit:

```
def illustrate(f, a):
    """
    Function to take a function and illustrate the limiting definition of a derivative at a given po
    """
    lst = []
    for h in srange(.01, 3, .01):
        lst.append([h,(f(a+h)-f(a))/h])
    return list_plot(lst, axes_labels=['$x$','$\\frac{f(%.02f+h)-f(%.02f)}{h}$' % (a,a)])
```

Figure 2: The derivative of $f(x) = 1 - x^2$ at $x = .5$ converging to -1 as $h \to 0$.

If we want to plot the tangent at a point α to a function we can use the following:

$$\begin{aligned}
y =& ax + b && \text{(definition of a straight line)}\\
& f'(a)x + b && \text{(definition of the derivative)}\\
& f'(a)x + f(a) - f'(a)a && \text{(we know that the line intersects } f \text{ at } (a, f(a)))
\end{aligned}$$

We can combine this with the approach of the previous piece of code to see how the tangential line converges as the limiting definition of the derivative converges:

```
def convergetangentialline(f, a, x1, x2, nbrofplots=50, epsilon=.1):
    """
    Function to make a tangential line converge
    """
    clrs = rainbow(nbrofplots)
    k = 0
    h = epsilon
    p = plot(f, x, x1, x2)
    while k < nbrofplots:
        tangent(x) = fdash(f, a, h) * x + f(a) - fdash(f, a, h) * a
        p += plot(tangent(x), x, x1, x2, color=clrs[k])
        h += epsilon
        k += 1
    return p
```

The plot shown in Figure 3 shows how the lines shown converge to the actual tangent to $1 - x^2$ as $x = 2$ (the red line is the 'closest' curve).

Figure 3: Lines converging to the tangent curve as $h \to 0$.

Note here that the last plot is given using the **real** definition of the derivative and not the approximation.

3 Conclusions

In this report I have explored the limiting definition of the limit showing how as $h \to 0$ we can visualise the derivative of a function. The code involved `https://sage.maths.cf.ac.uk/home/pub/18/` uses the differentiation capabilities of Sage but also the plotting abilities.

There are various other aspects that could be explored such as symbolic differentiation rules. For example:

$$\frac{dx^n}{dx} = (n+1)x^n \text{ if } x \neq -1$$

Furthermore it is interesting to not that there exists some functions that **are not** differentiable at a point such as the function $f(x) = \sin(1/x)$ which is not differentiable at $x = 0$. A plot of this function is shown in Figure 4.

Figure 4: None differentiable function at $x = 0$.