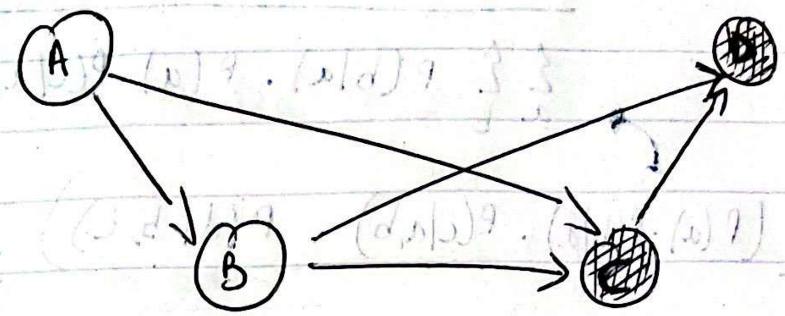
Hwk 6:

CSE 250A Andrew Ghafari A59020215.

6.1



a) $P(a,b|c,d)$

$$= \frac{P(a,b,c,d)}{P(c,d)} = \frac{\overset{1}{P(a)} \cdot \overset{2}{P(b|a)} \cdot \overset{3}{P(c|b,a)} \cdot \overset{4}{P(d|a,b,c)}}{\underset{5}{P(c)} \cdot \underset{6}{P(d|c)}}$$

1, 2, 3 in CPT ✓.

4. $P(d|a,b,c) = \underset{\text{in CPT}}{P(d|b,c)}$ $d \perp\!\!\!\perp a$ given $\{b,c\}$ d-sep.

5. $\sum_a \sum_b P(a,b,c) = \sum_a \sum_b P(a) \cdot P(b|a) \cdot P(c|b,a)$ - in CPTs

6. $P(d|c) = \sum_a \sum_b P(a,b,d|c)$

$$= \sum_a \sum_b \frac{P(a,b,c,d)}{P(c)}$$ ← This will cancel (5) in denom.

=> so we get

$$P(a,b \mid c,d) = \frac{P(a) \cdot P(b \mid a) \cdot P(c \mid a,b) \cdot P(d \mid b,c)}{\sum_a \sum_b P(b \mid a) \cdot P(a) \cdot P(c \mid a,b) \cdot P(d \mid b,c)}.$$

$$\sum_b \left(\sum_a \left(P(a) \cdot P(b \mid a) \cdot P(c \mid a,b) \cdot P(d \mid b,c) \right) \right).$$

b). $$P(a \mid c,d) = \sum_b \underbrace{P(a,b \mid c,d)}_{(a)}$$

- $$P(b \mid c,d) = \sum_a \underbrace{P(a,b \mid c,d)}_{(a)}$$

c) $$\mathcal{L} = \sum_t \log P(C = c_t, D = d_t)$$

$$= \sum_t \log \left(\sum_a \sum_b P(a,b,c_t,d_t) \right)$$

$$= \sum_t \log \left(\sum_a \sum_b \underbrace{P(a) \cdot P(b \mid a) \cdot P(c_t \mid a,b) \cdot P(d_t \mid b,c_t)}_{\text{just like in (a).}} \right)$$

in terms of decomposing the joint + conditional independence.

$d \perp\!\!\!\perp a$ given b,c.

d) As discussed in class:

A is a root node:

$$P(a) = \frac{1}{T} \sum_{t=1}^{T} P(a \mid c_t, d_t)$$

B has A as parent

$$P(b \mid a) = \frac{\sum_t P(b, a \mid c_t, d_t)}{\sum_t P(a \mid c_t, d_t)}$$

C has A, B as parents:

$$P(C=c \mid A=a, B=b) = \frac{\sum_t P(c, a, b \mid c_t, d_t)}{\sum_t P(a, b \mid c_t, d_t)}$$

$$= \frac{\sum_t P(c \mid a, b, c_t, d_t) \cdot P(a, b \mid c_t, d_t)}{\sum_t P(a, b \mid c_t, d_t)}$$

$$= \frac{\sum_t I(c, c_t) \cdot P(a, b \mid c_t, d_t)}{\sum_t P(a, b \mid c_t, d_t)}$$

D has B, C as parents.

$$P(d \mid b, c) = \frac{\sum_t P(d, b, c \mid d_t, c_t)}{\sum_t P(b, c \mid c_t, d_t)}$$

$$= \frac{\sum_t P(d \mid d_t, c_t, b, c) \cdot P(b \mid c_t, d_t) \cdot P(c \mid c_t, d_t, b)}{\sum_t P(c \mid c_t, d_t, b) \cdot P(b \mid c_t, d_t)}.$$

$$= \frac{\sum_t I(d, d_t) \cdot I(c, c_t) \cdot P(b \mid c_t, d_t)}{\sum_t I(c, c_t) \cdot P(b \mid c_t, d_t)}.$$

6.2

a) $\sum_{Z} P(Y=1, Z \mid X) = \sum_{Z} P(Z \mid X) \cdot P(Y=1 \mid Z, X)$

$= \sum_{Z} P(Z \mid X) \cdot (1 - P(Y=0 \mid Z, X))$.

$Y \perp\!\!\!\perp X$ Given Z disy.

$= \sum_{Z} P(Z \mid X) \cdot (1 - P(Y=0 \mid Z))$.

$= \sum_{Z} P(Z \mid X) \cdot \left(1 - \prod_{i=1}^{n} (I(Z_i, 0))\right)$.

$P(Z \mid X) = P(Z_1, \ldots Z_n \mid X_1 \ldots X_n)$

$= \prod P(Z_i \mid X_1 \ldots X_n) = \prod P(Z_i \mid X_i)$.

$P(Y=0 \mid X) = \sum_{Z \in \{0,1\}^n} P(Y=0, Z \mid X)$

$= \sum_{Z \in \{0,1\}^n} P(Z \mid X) \cdot P(Y=0 \mid Z, X)$ cond. indep. di.

$= \sum_{Z \in \{0,1\}^n} P(Z \mid X) \cdot P(Y=0 \mid Z)$

$= \sum_{Z \in \{0,1\}^n} \prod_{i=1}^{n} P(Z_i \mid X_i) \cdot \prod_{i=1}^{n} (I(Z_i, 0))$.

$= \prod_{i=1}^{n} P(Z_i = 0 \mid X_i)$

$= \prod_{i=1}^{n} (1 - p_i)^{X_i}$ (given).

So $P(Y=1 \mid X) = 1 - \prod_{i=1}^{n} (1 - p_i)^{X_i}$ H. Prova.

b) $\mathcal{L} = \frac{1}{T} \sum_t \log P\left(Y = y^{(t)} \mid X = \vec{x}^{(t)}\right)$

$x \in \{0,1\}^n \quad y \in \{0,1\}^n$

$P(Z_i = 1, X_i = 1 \mid X = x, Y = y)$

$$= \frac{P(Z_i = 1, X_i = 1, X = x, Y = y)}{P(X = x, Y = y)}$$

Bayes rule

$$= \frac{P(Y = y \mid Z_i = 1, X_i = 1, X = x) \cdot P(Z_i = 1, X_i = 1 \mid X = x)}{P(Y = y \mid X = x)}$$

$X \perp\!\!\!\perp Y$ given Z d-sep

$$= \frac{P(Y = y \mid Z_i = 1) \cdot P(X_i = 1 \mid X = x, Z_i = 1) \cdot P(Z_i = 1 \mid X = x)}{P(Y = y \mid X = x)}$$

$$= \frac{I(y, 1) \cdot I(X_i, 1) \cdot P(Z_i = 1 \mid X_i = x_i)}{P(Y = y \mid X = x)}.$$

$$= \begin{cases} 0 \text{ if } y = 0 \text{ or } x_i = 0 : \\ \dfrac{P(Z_i = 1 \mid X_i = 1)}{P(Y = y \mid X = x)} \text{ if } y = 1 \text{ \& } X_i = 1 \end{cases}$$

$$= \begin{cases} 0 & y=0 \text{ or } x_i=0 \\ \dfrac{p_i}{1 - \prod_j (1-p_j)^{x_j}} & \text{if } y=1 \text{ and } x_i=1 \end{cases}$$

$$= \frac{y \cdot x_i \cdot p_i}{1 - \prod_j (1-p_j)^{x_j}}$$ Proved.

c)

$$p_i \leftarrow \frac{\sum_t P(Z_i=1, X_i=1 \mid X=\vec{x}^{(t)}, Y=y^{(t)})}{\sum_t P(X_i=1 \mid X=\vec{x}^{(t)}, Y=y^{(t)})}$$

$$p_i \leftarrow \frac{\sum P(Z_i=1, X_i=1 \mid X=\vec{x}^{(t)}, Y=y^{(t)}}{\sum I(x_i^{(t)}, 1)}$$

$$p_i \leftarrow \frac{1}{T_i} \cdot \sum P(Z_i=1, X_i=1 \mid X=\vec{x}^{(t)}, Y=y^{(t)}$$

Scanned with CamScanner

6.3

a) $f(x) = \log \cosh(x)$

$$f'(x) = \frac{\sinh(x)}{\cosh(x)} = \tanh(x).$$

$$f'(x) = 0 \quad \tanh(x) = 0 \quad \frac{e^x - e^{-x}}{e^x + e^{-x}} = 0$$

$$e^x - e^{-x} = 0 \quad e^x = \frac{1}{e^x} \quad (e^x)^2 = 1$$

$$\underline{e^x = 1} \mid e^x = -1$$

D.N.E.

$$e^x = 1 \rightarrow \underline{\underline{x = 0}}.$$

minim only occurs iff. $x = 0$. Proved.

b) $$f''(x) = \frac{d}{dx} \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2}$$

$$= \frac{(e^x + e^{-x})^2}{(e^x + e^{-x})^2} - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} = 1 - \tanh^2(x)$$

$$= \operatorname{sech}^2(x).$$

range of sech(x) is $]0;1[$

$\text{sech}^2(x) < 1$ as well. Proud.

c). Python.

d) * $Q(x,x) = f(x) + \cancel{f'(x)(x-x)}^{\,0} + \cancel{\frac{1}{2}(x-x)^2}^{\,0}$

$= f(x)$. Proud.

* $Q(x,y) \geq f(x)$. ?

$$f(x) = f(y) + \int_y^x du \left(f'(y) + \int_y^u \underbrace{f''(v)\,dv}_{\leq 1} \right).$$

$$f(x) \leq f(y) + \int_y^x du \left(f'(y) + (u-y) \right)$$

$$\leq f(y) + f'(y) \cdot u \Big|_y^x + \frac{u^2}{2}\Big|_y^x - yu\Big|_y^x$$

$$\leq f(y) + f'(y) \cdot (x-y) + \frac{1}{2}(x^2-y^2) - yx + y^2$$

$$\leq f(y) + f'(y)(x-y) + \underbrace{\frac{1}{2}x^2 - yx + \frac{1}{2}y^2}$$

$$f(x) \leq f(y)\ f'(y)(x-y) + \frac{1}{2}(x-y)^2 \Rightarrow f(x) \leq Q(x,y)$$

Proud!

e) $x_{n+1} = \arg\min Q(x, x_n)$

$$= \arg\min \left(f(x_n) + f'(x_n)(x - x_n) + \frac{1}{2}(x - x_n)^2 \right)$$

$$= \arg\min \left(f(x_n) + f'(x_n)x - f'(x_n)x_n + \frac{1}{2}x^2 + \frac{1}{2}x_n^2 - x_n x_n \right)$$

$$= \arg\min \left(\frac{1}{2}x^2 + x\left(f'(x_n) - x_n\right) + \left(f(x_n) - f'(x_n)x_n + \frac{1}{2}x_n^2\right) \right)$$

$\Rightarrow ax^2 + bx + c.$

$$\arg\min = \frac{-b}{2a} = \frac{x_n - f'(x_n)}{2 \times \frac{1}{2}} = x_n - f'(x_n)$$

$$= \underline{x_n - \tanh(x)}.$$

CSE250Hw6

November 11, 2022

#Q2-d

[ ]:
```
import numpy as np
import warnings
import matplotlib.pyplot as plt
warnings.filterwarnings("ignore")
```

[ ]:
```
X = np.loadtxt("X.txt")
Y = np.loadtxt("Y.txt")
```

[ ]:
```
niter = 257
T = 267
n = 23

def likelihood(p, x, y):
  temp = np.prod((1-p)**x)
  return (1-y)*temp + y*(1-temp)

def conv(p, x, y):
  numi = p*y*x
  deno = 1-np.prod((1-p)**x)
  return numi/deno

Ti = []
for i in range(n):
  Ti.append(np.sum(X[:,i]))
```

[ ]:
```
printed = [2**i for i in range(9) ]
printed.insert(0,0)
def EM(X,Y):
  p = np.array([0.05]*n)
  Nmistakes = []
  LogL = []
  for i in range(niter):
    logl = 0
    nMis = 0
    eSum = 0
```

```
    for k in range(267):
      prob = likelihood(p,X[k],Y[k])
      logl += np.log(prob)
      eSum += conv(p,X[k],Y[k])
      if prob <= 0.5:
        nMis += 1
    p = eSum/Ti
    Nmistakes.append(nMis)
    logl /= 267
    LogL.append(logl)

    if i in printed:
      print(i, "\t" ,"\t", nMis, "\t" ,"\t" ,"\t" , logl )

  return Nmistakes,LogL
```

[ ]:
```
print("Iteration No ", "\t" , "Number of Mistakes " , "\t" , "LogLikelihood")
mLis, logLis = EM(X, Y)
```

```
Iteration No      Number of Mistakes      LogLikelihood
0                 175                     -0.9580854082157914
1                 56                      -0.49591639407753635
2                 43                      -0.40822081705839114
4                 42                      -0.3646149825001877
8                 44                      -0.3475006162087826
16                40                      -0.33461704895854844
32                37                      -0.32258140316749784
64                37                      -0.3148266983628559
128               36                      -0.3111558472151897
256               36                      -0.310161353474076
```

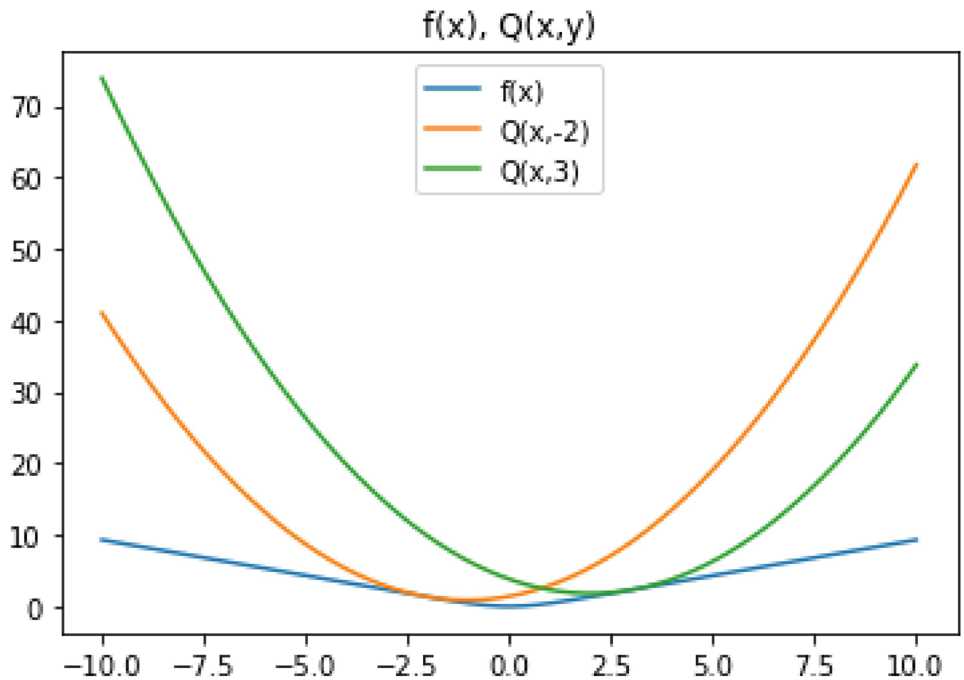
#Q3-c

[ ]:
```
def f(x):
    return np.log(np.cosh(x))

def df(x):
    return np.tanh(x)

def Q(x,y):
    return f(y) + df(y)*(x-y) + ((x-y)**2)/2
```

[ ]:
```
x = np.linspace(-10,10,10000)
plt.plot(x, f(x), label="f(x)")
plt.plot(x, Q(x,-2), label="Q(x,-2)")
plt.plot(x, Q(x,3), label="Q(x,3)")
```

```
plt.title('f(x), Q(x,y)')
plt.legend()
```

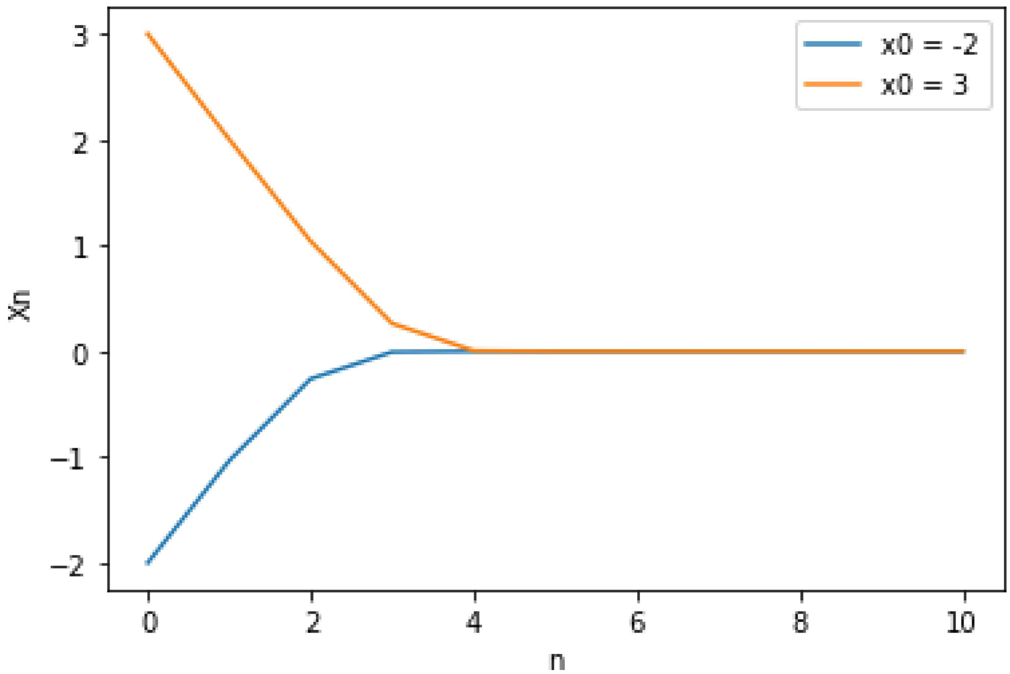
[ ]: <matplotlib.legend.Legend at 0x7fe9198eda90>



#Q3-f

[ ]:
```
def newUpdated(x0, n):
  xn = [x0]
  for i in range(len(n)-1):
      xn.append(xn[-1] - df(xn[-1]))
  return xn

n = np.arange(0, 11)
x1 = newUpdated(-2, n)
x2 = newUpdated(3, n)

plt.plot(n, x1, label="x0 = -2")
plt.plot(n, x2, label="x0 = 3")
plt.ylabel('Xn')
plt.xlabel('n')
plt.legend()
```

[ ]: <matplotlib.legend.Legend at 0x7fe9195ee750>


x0 = -2
x0 = 3
Xn
n
3
2
1
0
−1
−2
0
2
4
6
8
10