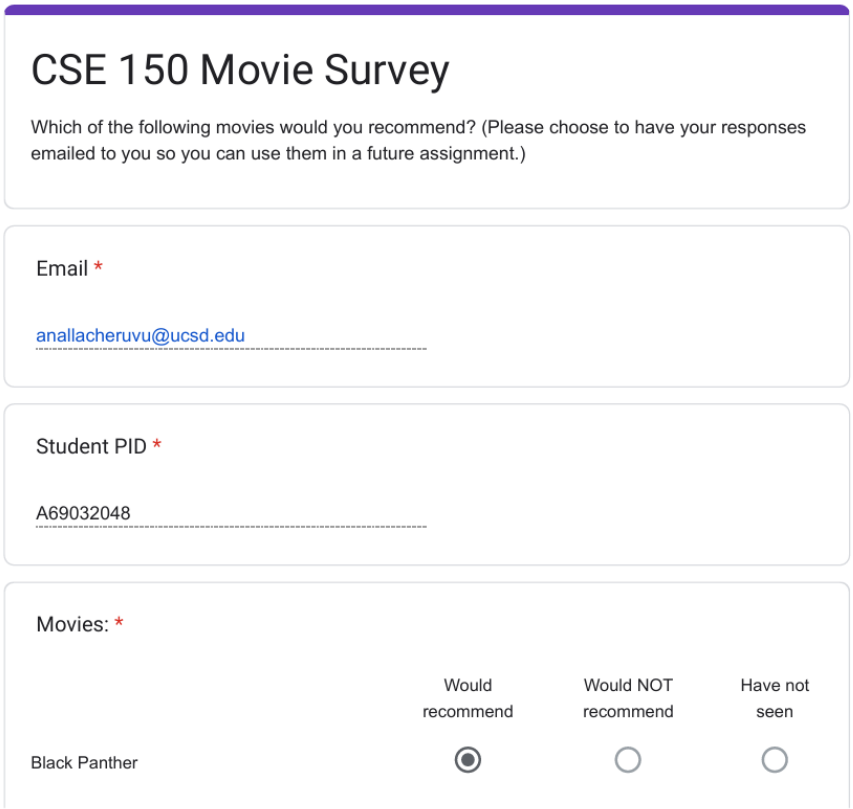
CSE 250A HW5

Akhil Nallacheruvu

November 2024

5.1 Survey



CSE 150 Movie Survey

Which of the following movies would you recommend? (Please choose to have your responses emailed to you so you can use them in a future assignment.)

Email *

anallacheruvu@ucsd.edu

Student PID *

A69032048

Movies: *

	Would recommend	Would NOT recommend	Have not seen
Black Panther	◉	○	○

Inception	◉	○	○
The Last Airbender	○	◉	○
The Hunger Games	◉	○	○
The Wolf of Wall Street	◉	○	○
World War Z	○	◉	○
Interstellar	○	◉	○
The Martian	◉	○	○
Iron Man 3	○	◉	○
La La Land	○	◉	○
Star Wars: The Last Jedi	○	◉	○
WALL-E	◉	○	○
The Dark Knight	◉	○	○
The Lord of the Rings: The Fellowship of the Ring	○	○	◉
The Matrix	◉	○	○
Star Trek Beyond	○	◉	○
Jurassic World	○	◉	○
Jurassic Park (1993)	◉	○	○
Deadpool 2	○	◉	○
Guardians of the Galaxy	○	◉	○
Mission: Impossible - Fallout	◉	○	○
Guardians of the Galaxy Vol. 2	○	◉	○

Zootopia	◉	○	○
Tron	◉	○	○
Star Wars: The Phantom Menace	○	○	◉
Moana	◉	○	○
Man of Steel	◉	○	○
Get Out	◉	○	○
Suicide Squad	◉	○	○
The Shape of Water	○	◉	○
The Avengers	◉	○	○
Mad Max: Fury Road	○	◉	○
Frozen	◉	○	○
The Imitation Game	◉	○	○
Ex Machina	○	○	◉
2001: A Space Odyssey	◉	○	○
Blade Runner 2049	○	○	◉
Terminator Genisys	○	○	◉
Terminator 2	○	○	◉
Avengers: Infinity War	◉	○	○
Coco	◉	○	○
Ant-Man and the Wasp	◉	○	○
Venom	◉	○	○

5.2 EM algorithm

a)

$$P(B=b|A=a) = \frac{\sum_{t=1}^{T} I(a,a_t) I(b,b_t)}{\sum_{t=1}^{T} I(a,a_t)}$$

$$P(C=c|A=a,B=b) = \frac{\sum_{t=1}^{T} I(a,a_t) I(b,b_t) I(c,c_t)}{\sum_{t=1}^{T} I(a,a_t) I(b,b_t)}$$

$$P(D=d|A=a,C=c) = \frac{\sum_{t=1}^{T} I(a,a_t) I(c,c_t) I(d,d_t)}{\sum_{t=1}^{T} I(a,a_t) I(c,c_t)}$$

b) $P(A=a,B=b|C=c,D=d) = \frac{P(A=a,B=b,C=c,D=d)}{P(C=c,D=d)}$

$= \frac{P(A=a)P(B=b|A=a)P(C=c|B=b,A=a)P(D=d|C=c,B=b,A=a)}{P(C=c)P(D=d|C=c)}$

$= \boxed{\frac{P(A=a)P(B=b|A=a)P(C=c|B=b,A=a)P(D=d|C=c,B=b)}{\Sigma_{A=a,B=b}P(A=a)P(B=b|A=a)P(C=c|B=b,A=a)P(D=d|C=c,B=b)}}$

c) $\boxed{P(a|c,d) = \Sigma_{B=b}P(A=a,B=b|C=c,D=d)}$

$\boxed{P(b|c,d) = \Sigma_{A=a}P(B=b,A=a|C=c,D=d)}$

d) $\mathrm{P(C=c}_t, D=d_t) = P(C=c_t)P(D=d_t|C=c_t) =$
$\Sigma_{A=a,B=b}P(A=a)P(B=b|A=a)P(C=c_t|B=b,A=a)P(D=d_t|C=c_t,B=b)$

$\boxed{\mathrm{L=}\Sigma_t log \Sigma_{A=a,B=b}P(A=a)P(B=b|A=a)P(C=c_t|B=b,A=a)P(D=d_t|C=c_t,B=b)}$

e)

$$P(A=a) \leftarrow \frac{\sum_{t=1}^{T} P(A=a|C=c_t, D=d_t)}{T}$$

$$P(B=b|A=a) \leftarrow \frac{\sum_{t=1}^{T} P(B=b, A=a|C=c_t, D=d_t)}{\sum_{t=1}^{T} P(A=a|C=c_t, D=d_t)}$$

$$P(C=c|A=a, B=b) \leftarrow \frac{\sum_{t=1}^{T} I(c, c_t) P(A=a, B=b|C=c_t, D=d_t)}{\sum_{t=1}^{T} P(A=a, B=b|C=c_t, D=d_t)}$$

$$P(D=d|B=b, C=c) \leftarrow \frac{\sum_{t=1}^{T} I(c, c_t) I(d, d_t) P(B=b|C=c_t, D=d_t)}{\sum_{t=1}^{T} I(c, c_t) P(B=b|C=c_t, D=d_t)}$$

5.3 EM algorithm for noisy-OR

a) P(Y=1—X) $= \Sigma_{Z \epsilon \{0,1\}^n} P(Y=1, Z|X) =$
$\Sigma_{Z \epsilon \{0,1\}^n} P(Y=1|Z, X) P(Z|X) = \Sigma_{Z \epsilon \{0,1\}^n} P(Y=1|Z) P(Z|X) =$
$\Sigma_{Z \epsilon \{0,1\}^n} \Pi_i^n P(Z_i|X_i) = 1 - \Pi_i^n P(Z_i=0|X_i) =$ $\boxed{\text{1-}\Pi_i^n (1-p_i)^{X_i}}$

$b) P(Z_i=1, X_i=1|X=x, Y=y) = \frac{\Sigma_{Z=\{0,1\}^n} P(Z_i=1, X_i=1, X=x, Z=z, Y=1)}{P(X=x, Y=y)} =$
$\frac{P(Z_i=1, X_i=1) P(X_i=1) \Sigma_{z_j} \Pi_j P(Z_j=z_j, X_j=x_j) P(Y=1|Z=z)}{P(Y=1|X=x)} =$ $\boxed{\frac{y x_i p_i}{1-\Pi_j (1-p_j)^{X_j}}}$

$c) p_i = P(Z_i=1|X_i=1)$
$P(Z_i=1|X_i=1) \leftarrow \frac{\Sigma_{t=1}^{T} P(Z_i=1, X_i=1|X=x^{(t)}, Y=y^{(t)})}{\Sigma_{t=1}^{T} P(X_i=1|X=x^{(t)}, Y=y^{(t)})} =$
$\frac{\Sigma_{t=1}^{T} P(Z_i=1, X_i=1|X=x^{(t)}, Y=y^{(t)})}{T_i}$

$\boxed{\text{p}_i \leftarrow \frac{\Sigma_{t=1}^{T} P(Z_i=1, X_i=1|X=x^{(t)}, Y=y^{(t)})}{T_i}}$

HW5

November 11, 2024

[1]:
```
import math
```

[3]:
```
with open('spectX-1.txt', 'r') as f:
    features = [list(map(int, line.rstrip(' \n').split(' '))) for line in f]

with open('spectY-1.txt', 'r') as f:
    labels = [int(line.rstrip("\n")) for line in f]
```

[5]:
```
num_samples = len(features)
iterations = 256
num_features = 23
probabilities = [0.05] * num_features
feature_counts = [0] * num_features
```

[7]:
```
for i in range(num_samples):
    for j in range(num_features):
        if features[i][j] == 1:
            feature_counts[j] += 1
```

[9]:
```
log_likelihood = [0] * (iterations + 1)
mistakes = [0] * (iterations + 1)
```

[11]:
```
for iteration in range(iterations + 1):
    for sample in range(num_samples):
        term_1 = 0
        term_2 = 1
        for feature in range(num_features):
            if features[sample][feature] == 1:
                term_1 += features[sample][feature] * (0 if␣
↪probabilities[feature] == 0 else math.log(1 - probabilities[feature]))
            term_2 *= (1 - probabilities[feature]) ** features[sample][feature]
        log_likelihood[iteration] += (1 - labels[sample]) * term_1 +␣
↪labels[sample] * math.log(1 - term_2)

    log_likelihood[iteration] /= num_samples
    mistakes[iteration] = 0
    for sample in range(num_samples):
        probability = 1
```

```
        for feature in range(num_features):
            probability *= (1 - probabilities[feature]) **␣
↪features[sample][feature]
        prob_of_1 = 1 - probability
        if (labels[sample] == 1 and prob_of_1 <= 0.5) or (labels[sample] == 0␣
↪and prob_of_1 >= 0.5):
            mistakes[iteration] += 1

    posteriors = []
    for sample in range(num_samples):
        denominator = 1
        for feature in range(num_features):
            denominator *= (1 - probabilities[feature]) **␣
↪features[sample][feature]
        posteriors.append([labels[sample] * features[sample][feature] *␣
↪probabilities[feature] / (1 - denominator) for feature in range(num_features)])

    for feature in range(num_features):
        probabilities[feature] = sum(posteriors[sample][feature] for sample in␣
↪range(num_samples)) / feature_counts[feature]
```

[12]:
```
for iteration in [0, 1, 2, 4, 8, 16, 32, 64, 128, 256]:
    print(f"Iteration: {iteration} \t Mistakes: {mistakes[iteration]} \t Log␣
↪Likelihood: {round(log_likelihood[iteration], 5)}")
```

```
Iteration: 0     Mistakes: 175   Log Likelihood: -0.95809
Iteration: 1     Mistakes: 56    Log Likelihood: -0.49592
Iteration: 2     Mistakes: 43    Log Likelihood: -0.40822
Iteration: 4     Mistakes: 42    Log Likelihood: -0.36461
Iteration: 8     Mistakes: 44    Log Likelihood: -0.3475
Iteration: 16    Mistakes: 40    Log Likelihood: -0.33462
Iteration: 32    Mistakes: 37    Log Likelihood: -0.32258
Iteration: 64    Mistakes: 37    Log Likelihood: -0.31483
Iteration: 128   Mistakes: 36    Log Likelihood: -0.31116
Iteration: 256   Mistakes: 36    Log Likelihood: -0.31016
```