Matlab Problem of the First Practice Day

Maximum Likelihood, Maximum a Posterior, and Bayesian Learning

경마에서 어떤 말이 이길 지를 예측하는 문제를 생각해보자. 두 경마 전문가 (모델, θ)이 있는데, 한 전문가는 베테랑 전문가 ($\theta 1$)이고, 다른 전문가는 신참 전문가 ($\theta 2$)이다. 당신은 베테랑 전문가를 신참 전문가보다 4배 더 신뢰해왔다. 경마에는 두 말이 우승을 다투는데 1번말 ($h1$)과 2번말 ($h2$)이 있다.

경마장에는 매일 경기가 열린다. 전문가들은 말이 우승할 확률은 매 경기마다 독립이라고 생각한다. 그런데, 베테랑 전문가는 매 경기에 대하여 2번말이 우승할 가능성이 60%는 된다고 주장한다. 한편, 신참 전문가는 매 경기에 대하여 2번말이 우승할 가능성이 30%밖에 안된다고 주장했다. 그러나 이미 첫째 날 (x_1)과 둘째 날 (x_2)의 경마 경기가 열렸고, 두 번 모두 1번말이 이겼다.

당신은 전문가들의 주장 (모델들의 후보)과 현재까지의 경기 결과 (데이터)를 가지고 다음 날 경마장에서 돈을 걸고 싶다. 그런데, 셋째 날 (x_3) 2번말이 우승할 확률은 몇 %라고 생각하면 좋을까? (그런데, 이 문제는 감독학습일까? 천천히 생각해보자.)

각 전문가가 생각하는 2번말의 우승 확률은 다음과 같다.

$$p(x = h2 \mid \theta = \theta 1) = 0.6$$

$$p(x = h2 \mid \theta = \theta 2) = 0.3$$

그렇다면, 첫째 날과 둘째 날의 결과에 대해 각 전문가가 기대했던 확률 (likelihood)은 다음과 같을 것이다. (여기에 무슨 가정이 쓰였는가?)

$$p(x_1 = h1, x_2 = h1 \mid \theta = \theta 1) = 0.16$$

$$p(x_1 = h1, x_2 = h1 \mid \theta = \theta 2) = 0.49$$

모델이 주어졌을 때, 데이터의 확률이 어느 정도인지에 대해 생각하는 것은 쉽다. 하지만, 모델의 확률에 대해 생각하는 것은 쉽지 않다. 예컨대, Gaussian 분포가 주어졌을 때, 각 point의 확률

에 대해 생각하는 것은 쉽다. 하지만, Gaussian 분포의 표준편차가 0.4일 확률에 대해 생각하는 것은 직관적이지 않다. 하지만, 지금은 모델이 가질 수 있는 값의 경우의 수가 두 가지라고 가정하자. (Gaussian 분포에서 평균이나 표준편차가 가질 수 있는 값의 경우의 수는 몇 가지인가?) 위의 진술에 따르면 각 모델의 확률 (prior)은 다음과 같다.

$$p(\theta = \theta 1) = 0.8$$

$$p(\theta = \theta 2) = 0.2$$

우리가 풀고 싶은 문제는 다음 확률을 구하는 것이다. (정말 이 진술이 맞을까? θ 는 왜 없을까?) 이 확률을 구해야 어느 배당률에서부터 배팅하는 게 좋을 지를 결정할 수 있을 것이다.

$$P(x_3 = h2 \mid x_1 = h1, x_2 = h1)$$

수업 시간에 배웠던 바를 떠올려보자. 이 문제에서 ML과 MAP 그리고 Bayesian 방법은 다음과 같이 적용된다.

1. ML

Maximum Likelihood 방법에서는 likelihood $p(x_1, x_2 \mid \theta)$ 를 최대화하는 모델 θ를 선택한다.

Learning	$\theta_{ML} = \arg\max_{\theta} p(x_1, x_2 \mid \theta)$
Inference	$p(x_3 \mid x_1, x_2) \approx p(x_3 \mid \theta_{ML})$

2. MAP

Maximum a Posterior 방법에서는 Posterior $p(\theta \mid x_1, x_2)$ 를 최대화하는 모델 θ를 선택한다.

Learning	$\theta_{MAP} = \arg\max_{\theta} p(\theta \mid x_1, x_2)$ $= \arg\max_{\theta} p(x_1, x_2 \mid \theta) p(\theta)$
Inference	$p(x_3 \mid x_1, x_2) \approx p(x_3 \mid \theta_{MAP})$

3. Bayesian

Bayesian Learning에서는 Posterior $p(\theta | x_1, x_2)$ 를 추정한다.

Learning	$p(\theta \| x_1, x_2) = \frac{p(x_1, x_2 \| \theta) p(\theta)}{p(x_1, x_2)} = \frac{p(x_1, x_2 \| \theta) p(\theta)}{\sum_{\theta} p(x_1, x_2 \| \theta) p(\theta)}$
Inference	$p(x_3 \| x_1, x_2) = \sum_{\theta} p(x_3 \| \theta) p(\theta \| x_1, x_2)$

(x3의 조건부 확률을 구하기 위하여 수식 전개에서 어떤 가정이 사용되었는가? Sum rule, Product rule, Bayesian rule이 어디에 쓰였는지 생각해보자.)

다음은 Maximum Likelihood 학습 및 추론을 Matlab Code로 구현한 것이다.

```
% ptheta: model의 확률 p(θ), 1*2 vector
ptheta = [0.8 0.2];
% px_theta: model이 주어졌을 때에 데이터 x의 조건부 확률 p(x|θ), 2*2 matrix
px_theta = [0.4 0.6 ; 0.7 0.3 ];
% x12: 첫째날 ,둘째날의 우승 말 번호
x12 = [1 1];
% x3_predict: 셋째날에 맞추고자 하는 말의 번호
x3_predict = 2;

% px12_theta: p(x1=h1, x2=h1|θ)
px12_theta = px_theta(:,x12(1))' .* px_theta(:,x12(2))';

[max_val max_idx] = max(px12_theta);
answer_ml = px_theta(max_idx, x3_predict)
```

Matlab Code: ML Learning and Inference

이를 바탕으로, MAP이나 Bayesian 방식으로 학습 및 추론을 하는 Matlab Code를 만들고, 이 경우 $P(x_3 = h2 | x_1 = h1, x_2 = h1)$ 값이 얼마가 나오는 지 확인해보자.

```
px12theta = px12_theta .* ptheta;

[max_val max_idx] = max(px12theta);
answer_map = px_theta(max_idx,x3_predict)


px_12theta = px12theta/sum(px12theta);
answer_bayesian = px_theta(:,x3_predict)' * px_12theta'
```

Matlab Code: MAP, and Bayesian Learning

APPENDIX

1. 본 문서의 맥락에서 Prior, Likelihood, Posterior, Evidence

Prior	$p(\theta)$
Likelihood	$p(x_1, x_2 \mid \theta)$
Posterior	$p(\theta \mid x_1, x_2)$
Evidence	$p(x_1, x_2)$

2. Sum rule, Product rule, Bayes rule

Sum rule	$p(x) = \sum_Y p(X, Y)$
Product rule	$p(X,Y) = p(X \mid Y)p(Y)$ $= p(Y \mid X)p(X)$
Bayes rule	$p(X \mid Y) = \dfrac{p(Y \mid X)p(X)}{p(Y)}$