Q1:Bayesian Inference

(a)

F represents Fair coin; H represents Head.

$$P(F) = 0.5; P(\neg F) = 0.5; P(H|F) = 0.5; P(H|\neg F) = 0.5$$

Therefore

$$P(F|H) = \frac{P(F \cdot H)}{P(H)} = \frac{P(H|F)P(F)}{P(H|F)P(F) + P(\neg F)P(H|\neg F)}$$
$$= \frac{(0.5 \times 0.5)}{(0.5 \times 0.5 + 0.5 \times 0.8)} \approx 0.38$$

(b)

Now, the probability of the coin being fair is updated to 0.38. Then, recompute the posterior probability of the coin being fair.

$$P(F|H) = \frac{P(F \cdot H)}{P(H)} = \frac{P(H|F)P(F)}{P(H|F)P(F) + P(\neg F)P(H|\neg F)}$$
$$= \frac{(0.5 \times 0.38)}{(0.5 \times 0.38 + (1 - 0.38) \times 0.8)} \approx 0.28$$

(c)

In [13]:

```
Pr_F=0.28
Pr_Not_F=1-Pr_F
Pr_H_given_F=0.5
Pr_H_given_not_F=0.8
n=0
while Pr_F>=0.05:
    Pr_F=(Pr_H_given_F*Pr_F)/(Pr_H_given_F*Pr_F+Pr_Not_F*Pr_H_given_not_F)
    Pr_Not_F=1-Pr_F
    n=n+1
print(n)
```

5

Therefore, after 5 more flips, my belief about the probability of the coin being fair drops to below 0.05.

Q2: Fun with Linear Regression

(a)(1)

Because

$$\epsilon \sim N(0, \sigma)$$

Then, we can figure out

$$y^{(i)}|x^{(i)} \sim N(\omega^T x^{(i)}, \sigma)$$

Compute

$$L(\omega) := \prod_{i=1}^{n} P(y^{(i)}|x^{(i)}, \sigma, \omega)$$
$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y^{(i)}-\omega^T x^{(i)})^2}{2\sigma^2}}$$

(a)(2)

$$LL(\omega) := lnL(\omega)$$
$$= ln(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y^{(i)}-\omega^T x^{(i)})^2}{2\sigma^2}})$$
$$= \sum_{i=1}^{n} ln(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y^{(i)}-\omega^T x^{(i)})^2}{2\sigma^2}})$$
$$= \sum_{i=1}^{n} ln(\frac{1}{\sqrt{2\pi\sigma^2}}) + \sum_{i=1}^{n} ln(e^{\frac{-(y^{(i)}-\omega^T x^{(i)})^2}{2\sigma^2}})$$
$$= nln(\frac{1}{\sqrt{2\pi\sigma^2}}) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y^{(i)} - \omega^T x^{(i)})^2$$

Remove constant and terms that don't have ω

Therefore,

$$\hat{\omega} = \underset{\omega}{argmax} - \sum_{i=1}^{n} (y^{(i)} - \omega^T x^{(i)})^2$$
$$= \underset{\omega}{argmin} \sum_{i=1}^{n} (\omega^T x^{(i)} - y^{(i)})^2$$

When X represents a ($n \times p$) matrix which has n data samples and p features, Y represents a ($n \times 1$) matrix which has n results, $\min_{\omega} ||X\omega - Y||_2^2$ yields the maximizer of the posterior.

(b)(1)

Because

$$\epsilon \sim N(0, \sigma)$$
$$\omega \sim N(0, \tau)$$

Then, we can compute posterier by following steps:

$$M(\omega) := \prod_{i=1}^{n} P(\omega|y^{(i)}, x^{(i)})$$
$$= \prod_{i=1}^{n} \frac{P(y^{(i)}, x^{(i)}|\omega)g(\omega)}{h(y^{(i)}, x^{(i)})}$$
$$= \prod_{i=1}^{n} \frac{P(y^{(i)}|x^{(i)}, \omega)P(x^{(i)})g(\omega)}{h(y^{(i)}, x^{(i)})}$$

Remove constant multipliers that don't have ω

$$M(\omega) = \prod_{i=1}^{n} P(y^{(i)}|x^{(i)}, \omega)g(\omega)$$
$$= \prod_{i=1}^{n} L(\omega)g(\omega)$$
$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y^{(i)}-\omega^T x^{(i)})^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\tau^2}} e^{\frac{-\omega^2}{2\tau^2}}$$

(b)(2)

Compute the log-likelihood of the posterier:

$$m(\omega) := lnM(\omega)$$
$$= ln(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y^{(i)}-\omega^T x^{(i)})^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\tau^2}} e^{\frac{-\omega^2}{2\tau^2}})$$
$$= \sum_{i=1}^{n} ln(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y^{(i)}-\omega^T x^{(i)})^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\tau^2}} e^{\frac{-\omega^2}{2\tau^2}})$$
$$= \sum_{i=1}^{n} ln(\frac{1}{\sqrt{2\pi\sigma^2}}) + \sum_{i=1}^{n} ln(e^{\frac{-(y^{(i)}-\omega^T x^{(i)})^2}{2\sigma^2}}) + \sum_{i=1}^{n} ln(\frac{1}{\sqrt{2\pi\tau^2}}) + \sum_{i=1}^{n} ln(e^{\frac{-\omega^2}{2\tau^2}})$$
$$= nln(\frac{1}{\sqrt{2\pi\sigma^2}}) + nln(\frac{1}{\sqrt{2\pi\tau^2}}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y^{(i)} - \omega^T x^{(i)})^2 - \frac{n}{2\tau^2}\omega^2$$

Remove constant and terms that don't have ω

Therefore,

$$\omega_{MAP} = \underset{\omega}{argmax} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y^{(i)} - \omega^T x^{(i)})^2 - \frac{n}{2\tau^2}\omega^2$$
$$= \underset{\omega}{argmin}\ \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y^{(i)} - \omega^T x^{(i)})^2 + \frac{n}{2\tau^2}\omega^2$$
$$= \underset{\omega}{argmin}\sum_{i=1}^{n}(y^{(i)} - \omega^T x^{(i)})^2 + \frac{n\sigma^2}{\tau^2}\omega^2$$

When X represents a ($n \times p$) matrix which has n data samples and p features, Y represents a ($n \times 1$) matrix which has n results, $\lambda = \frac{n\sigma^2}{\tau^2}$, $\min_{\omega} ||X\omega - Y||_2^2 + \lambda||\omega||_2^2$ yields the maximizer of the posterior.