```
out = 1
if y == 1:
    out *= θ
else:
    out *= 1 - θ
```

lax.conditional

$\theta^{x}(1-\theta)^{1-x}$

OR

$\theta x + (1-\theta)(1-x)$

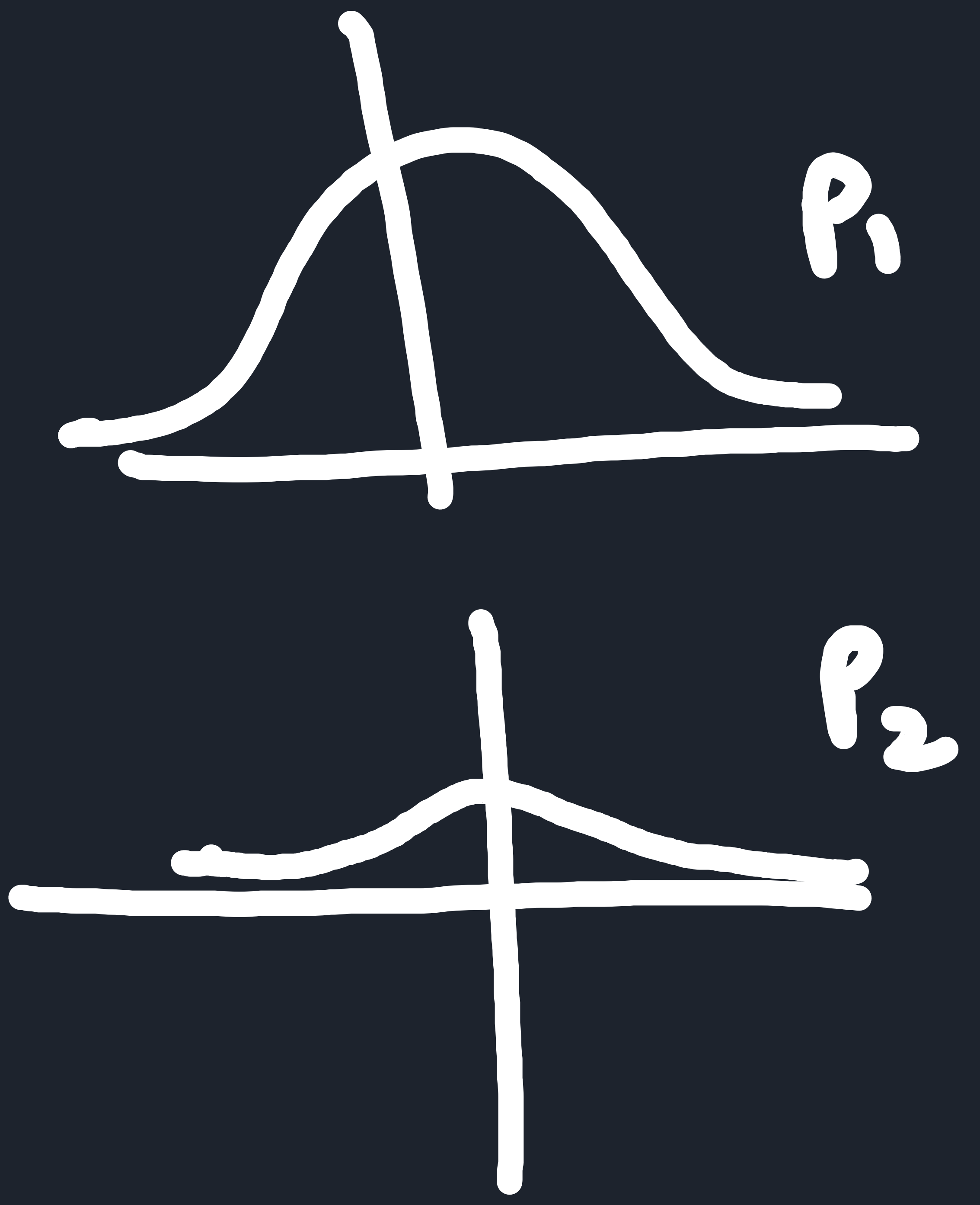
$$P(\theta|D) = \frac{P(D|\theta)\cdot P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, P(D|\theta)$$

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\, P(\theta|D)$$

$P_1(x) = N(0,1)$

$P_2(x) = \frac{N(0,1)}{40}$



$$\boxed{\begin{aligned} &\arg\max_x P_1(x) = 0 \\ &= \arg\max_x P_2(x) \\ &\text{MODE} \end{aligned}}$$

$$P(\theta|D) \propto P(D|\theta).P(\theta)$$

$$\therefore \hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\, P(D|\theta).P(\theta)$$

$$= \underset{\theta}{\operatorname{argmax}}\, P(D,\theta)$$

Lin-reg $y_i \sim N(x_i^T\theta, \sigma^2)$

Log. $y_i \sim Bern(\sigma(x_i^T\theta))$

Prior

$P(\theta) = ?$

$P(\theta) = N(0, b^2 I)$

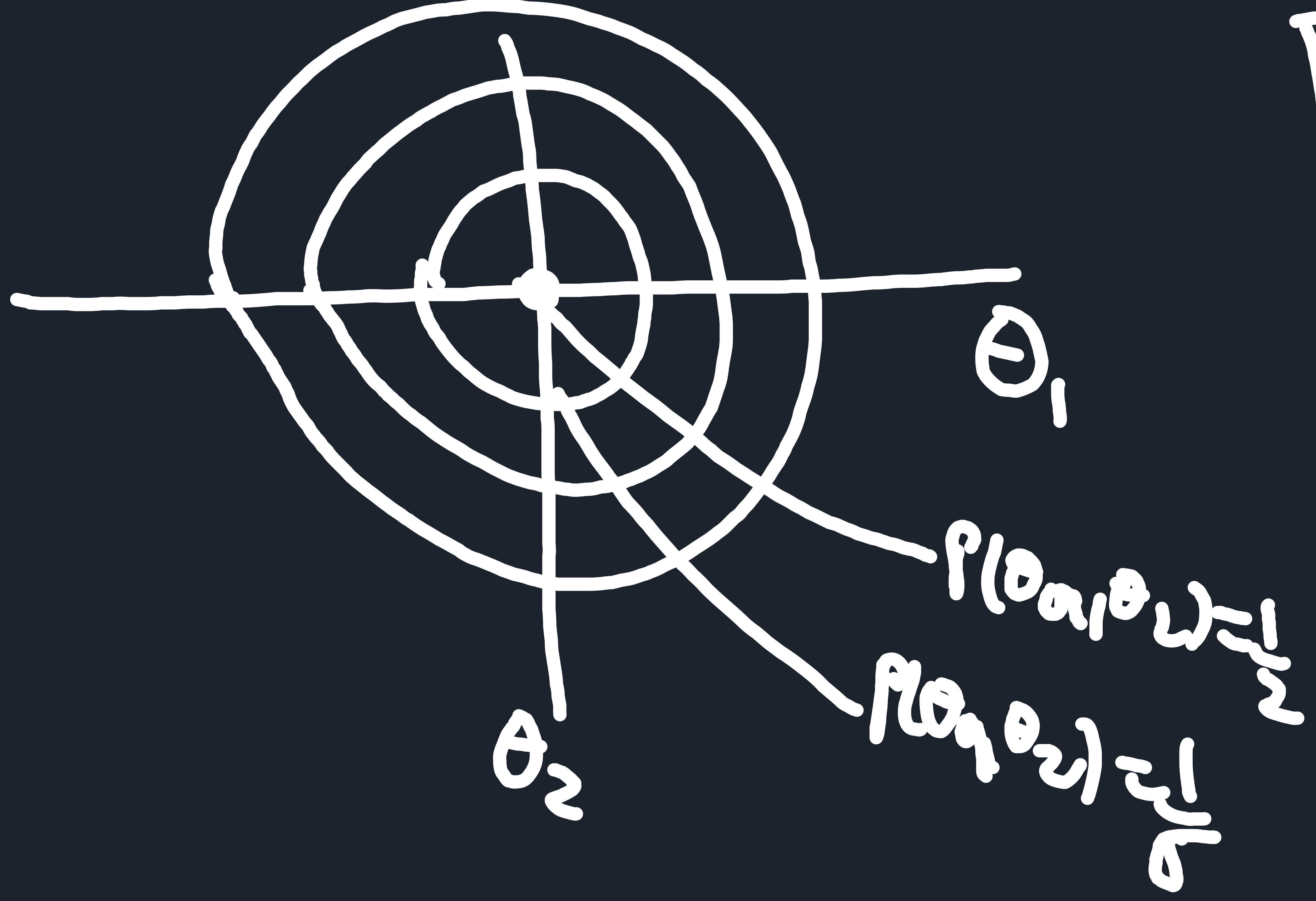
$= N\left(\theta \,\middle|\, \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, b^2 \begin{bmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ & & \ddots \\ & & & 1 \end{bmatrix}\right)$

MVNormal

$P(\theta \mid \mu, \Sigma) = \frac{1}{(2\pi|\Sigma|)^{1/2}} \exp\left\{-\frac{1}{2}(\theta-\mu)^T \Sigma^{-1} (\theta-\mu)\right\}$

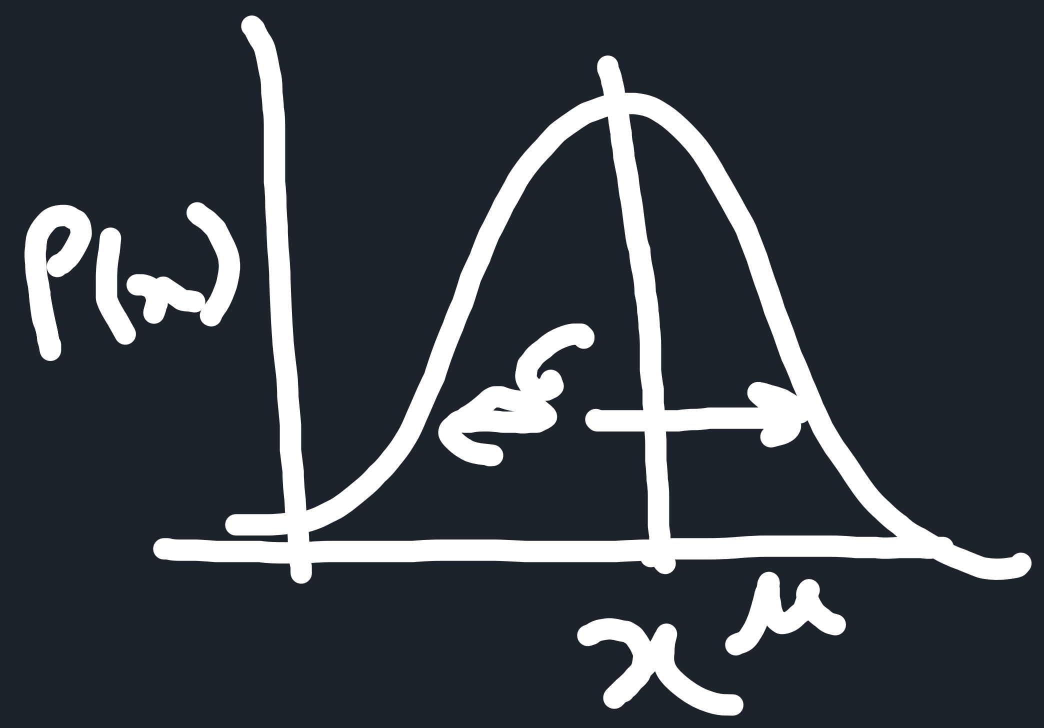
$b=1$

$$P(\theta) = N\left(\theta \,\middle|\, \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$$

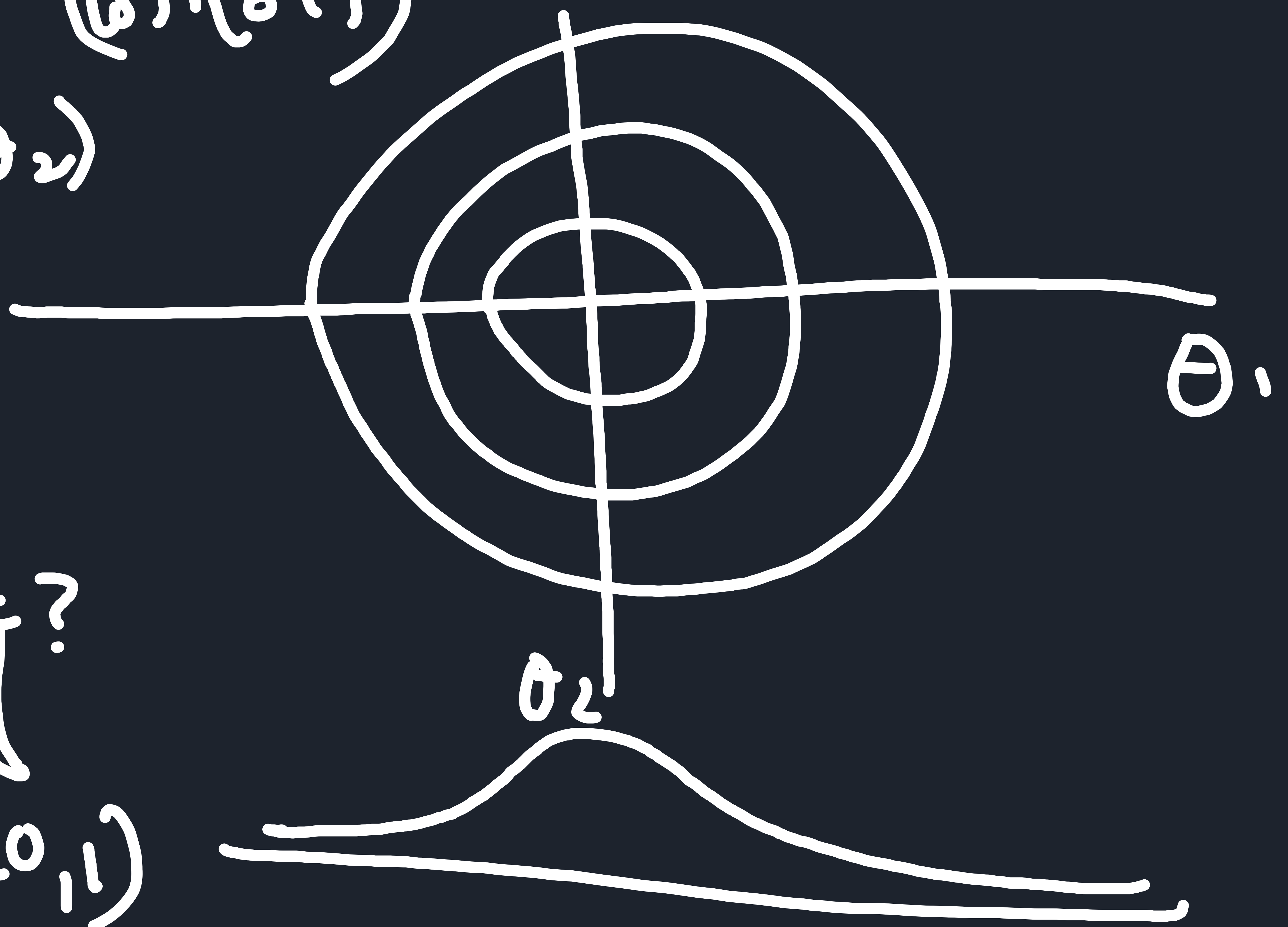


Aside

$P(x) = N(\mu, \sigma^2)$



$$p(\theta_1, \theta_2) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

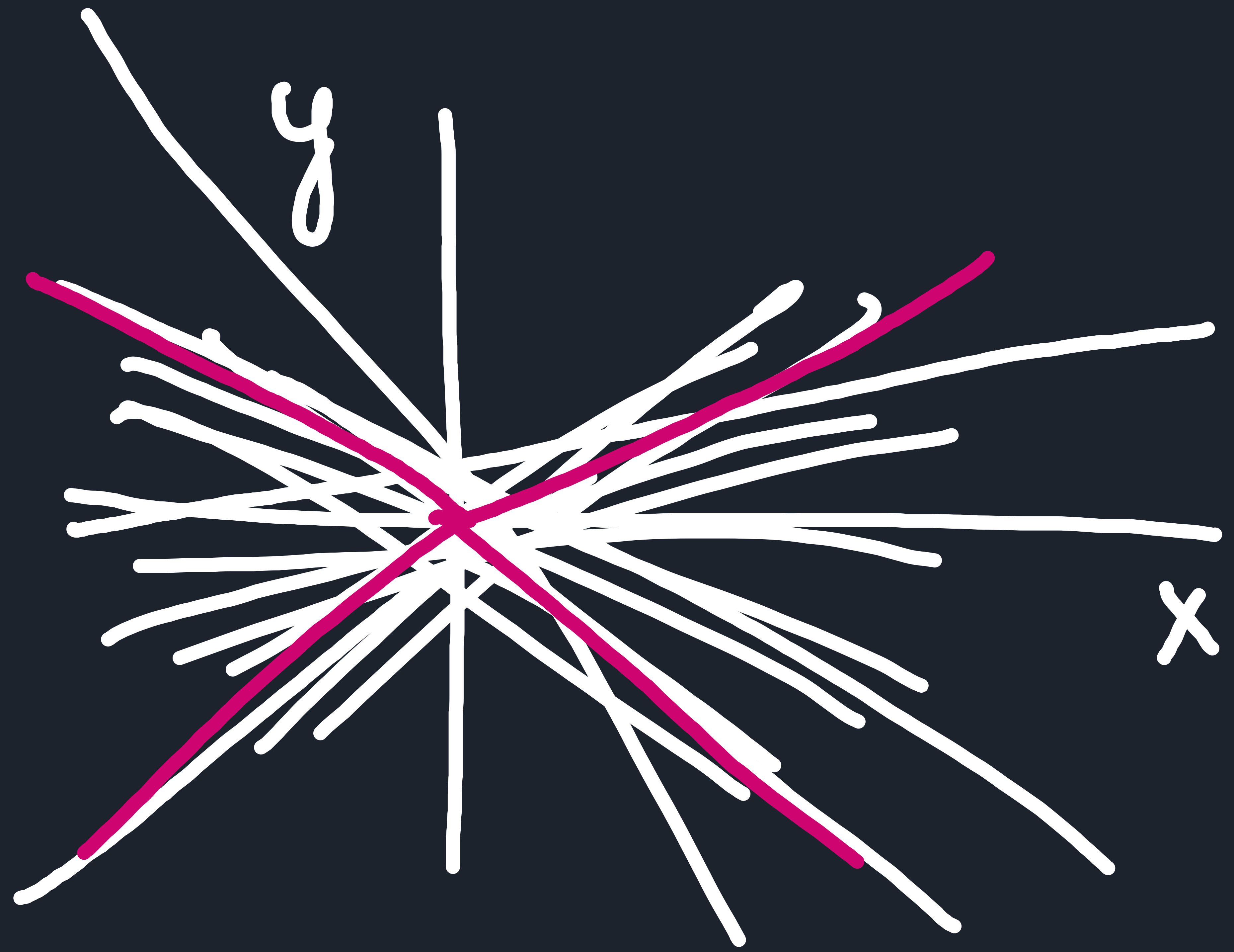


$$p(\theta_1) = ?$$

$$\mathcal{N}(0,1)$$

Prior

$$P(\theta) = N\left(\begin{bmatrix}0\\0\end{bmatrix}, b^2\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$$



$$y = \theta_0 + \theta_1 x$$

Linear Reg.

$$P(\theta) = N(0, b^2 I)$$

$$\hat{\theta}_{MAP} = ? = \underset{\theta}{\operatorname{argmin}} J(\theta)$$

$$= \underset{\theta}{\operatorname{argmax}} \left\{ P(D|\theta) \cdot P(\theta) \right\}$$

$$= \underset{\theta}{\operatorname{argmax}} \left\{ \log P(D|\theta) + \log P(\theta) \right\}$$

$$P(\theta) \propto \exp\left(-\frac{1}{2}(\theta - 0)^T \Sigma^{-1} (\theta - 0)\right)$$

$$\log P(\theta) \propto -\frac{1}{2} \theta^T \Sigma^{-1} \theta$$

$$\boxed{\Sigma = b^2 I \Rightarrow \Sigma^{-1} = \frac{1}{b^2} I}$$

$$\therefore \log P(\theta) \propto -\frac{1}{2b^2} \theta^T I \theta$$

$$\boxed{\log P(\theta) \propto -\frac{1}{2b^2} \theta^T \theta}$$

$$\log \text{Prior} = \log P(\theta) = \frac{-1}{2b^2}\theta^T\theta$$

$$\log\text{.likelihood} = \frac{-1}{2\sigma^2}(y - X\theta)^T(y - X\theta)$$

$$\therefore \boxed{\begin{aligned}\hat{\theta}_{MAP} &= \underset{\theta}{\operatorname{argmin}}\left(N.L.L. + N.L.\text{Prior}\right)\\ &= \underset{\theta}{\operatorname{argmin}}\ \frac{(y - X\theta)^T(y - X\theta)}{2\sigma^2} + \frac{\theta^T\theta}{2b^2}\end{aligned}}$$

$$J(\theta) = (y - X\theta)^T(y - X\theta) + \lambda\theta^T\theta$$

$$\theta^T\theta$$

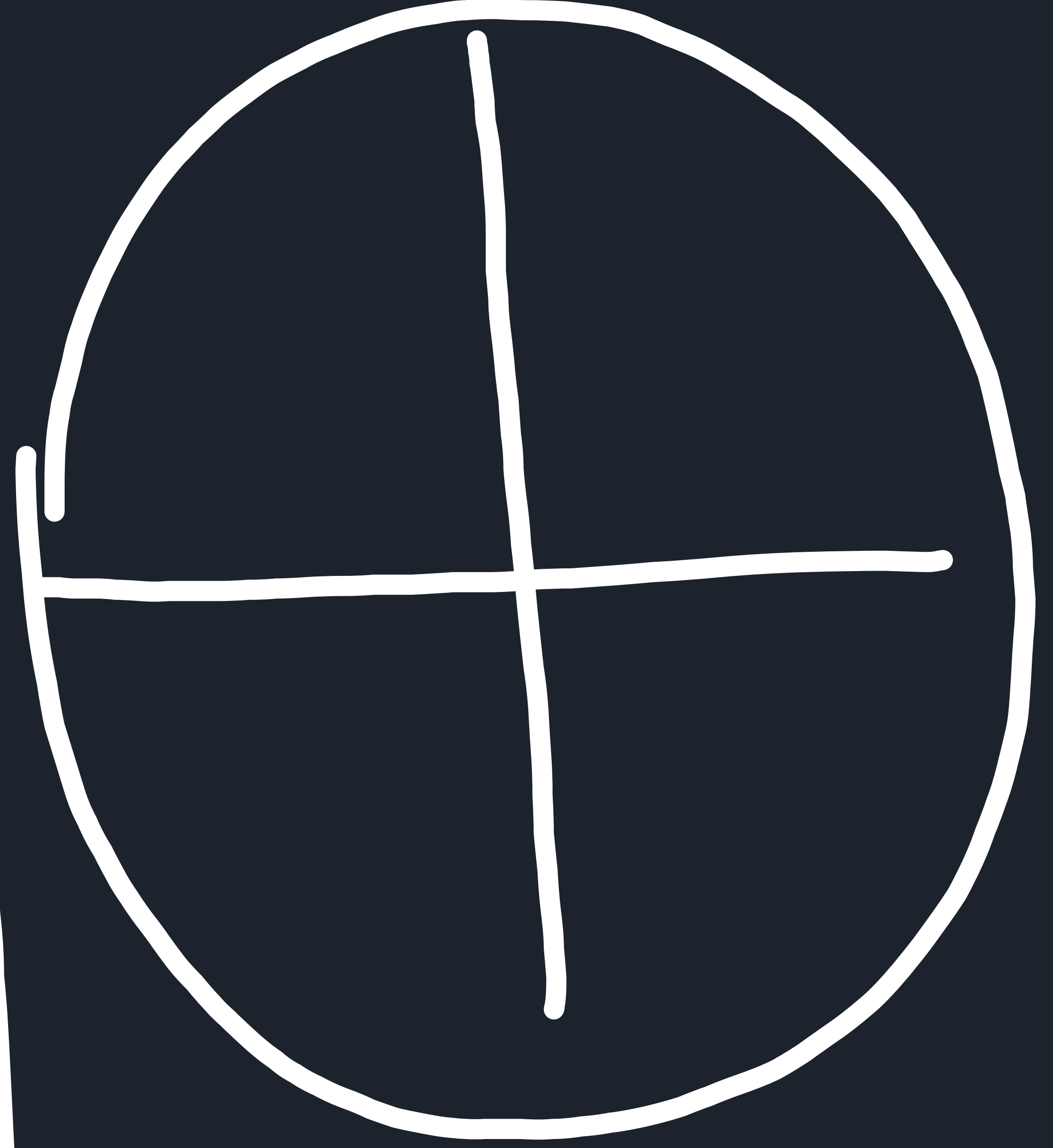
$$\lambda = \frac{\sigma^2}{b^2}$$

$b^2 I$

$b = 1$

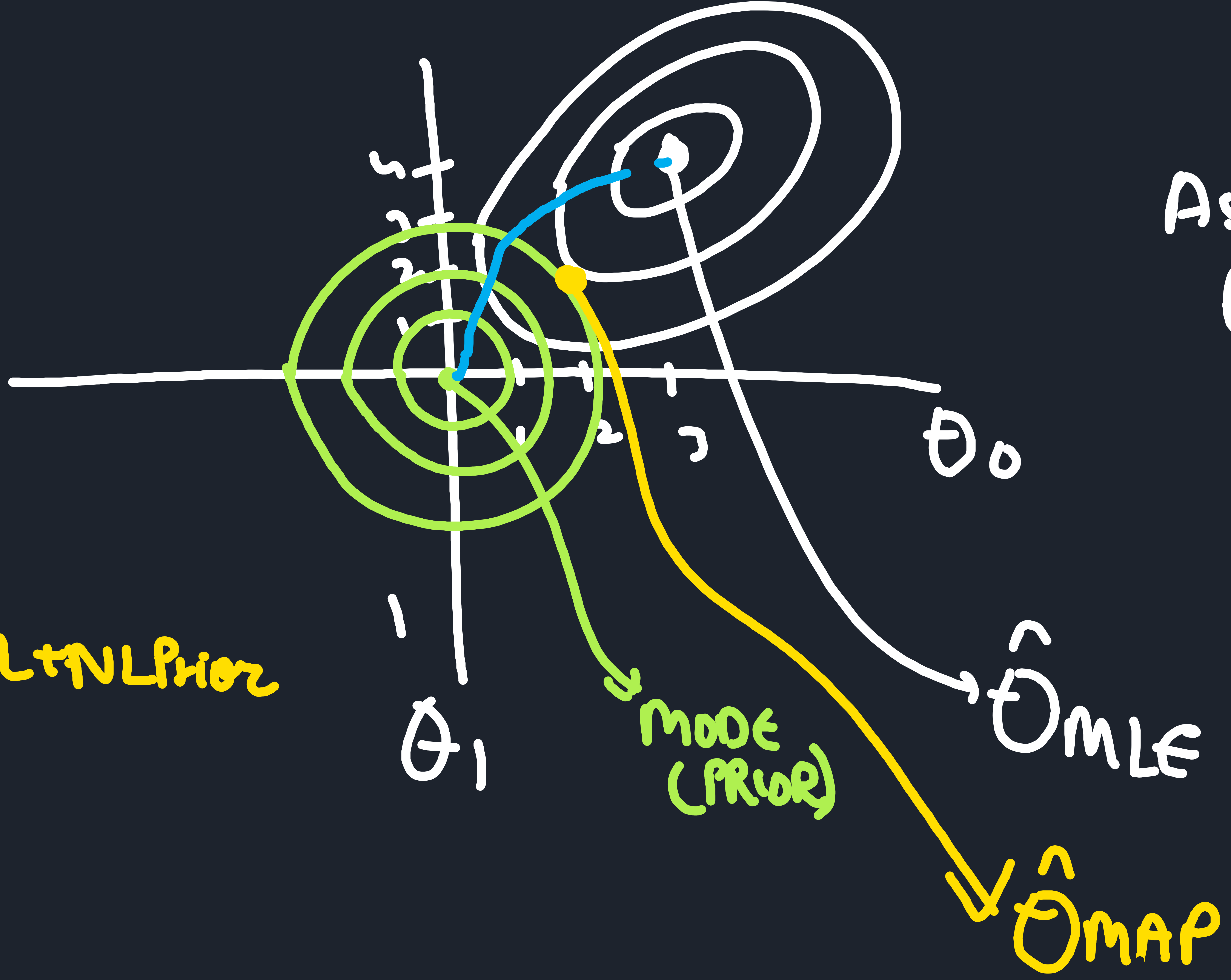
$b = 0.001$

$$N\left(\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}, \begin{bmatrix}.001 & & \\ & \ddots & \\ & & .001\end{bmatrix}\right)$$



$b = 1000$

Assume $\theta_{TRUE} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$



$J(\theta) = NLL + NLPrior$

$$\Theta \sim N(\theta \mid \mu, \sigma^2 I)$$

$$b \sim N(\mu, \sigma)$$

$P(\theta)$

Dataset D_1

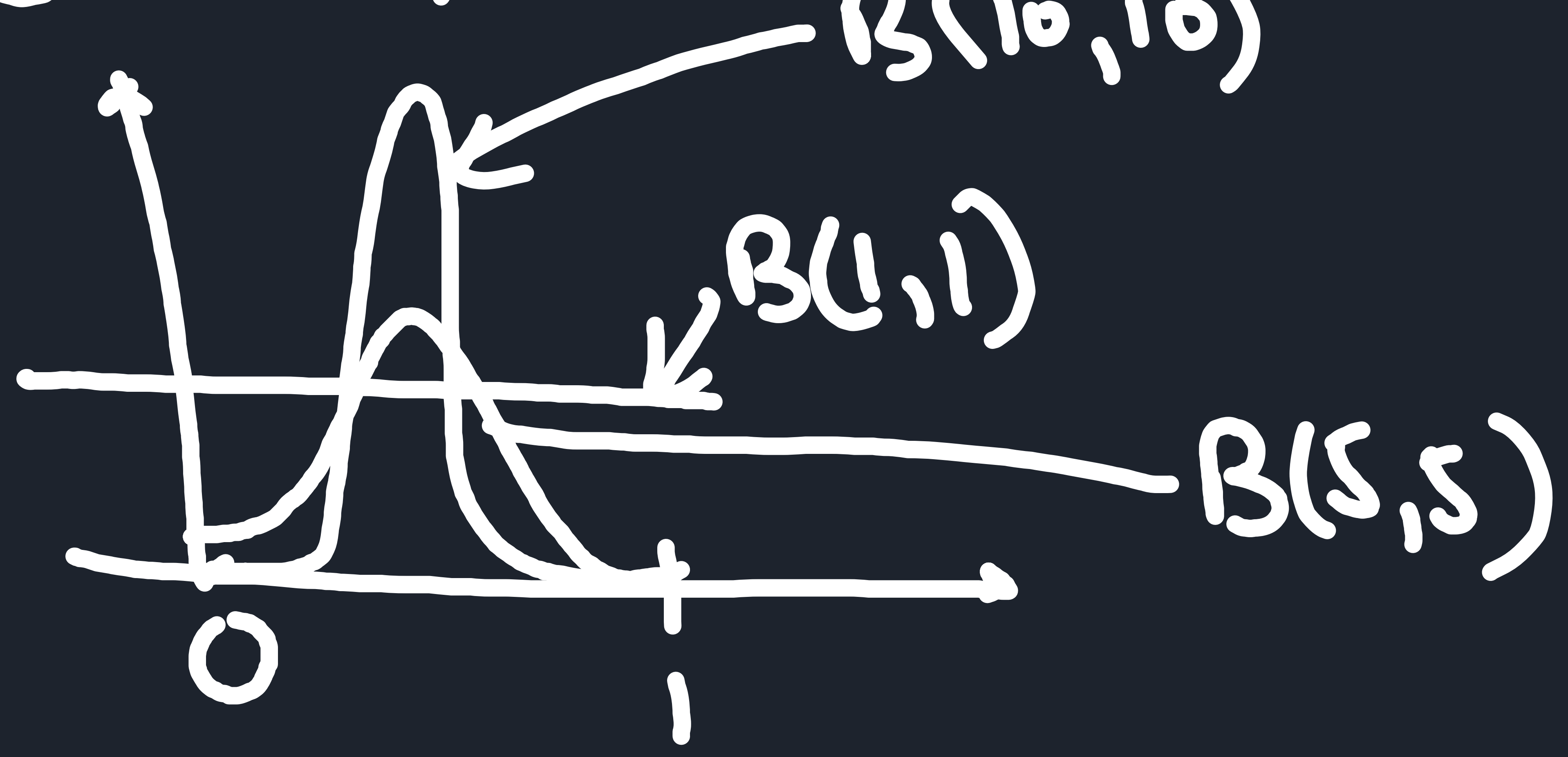
$P(\theta \mid D_1)$: Prior

Data : D_2

$P(\theta \mid D_1, D_2)$

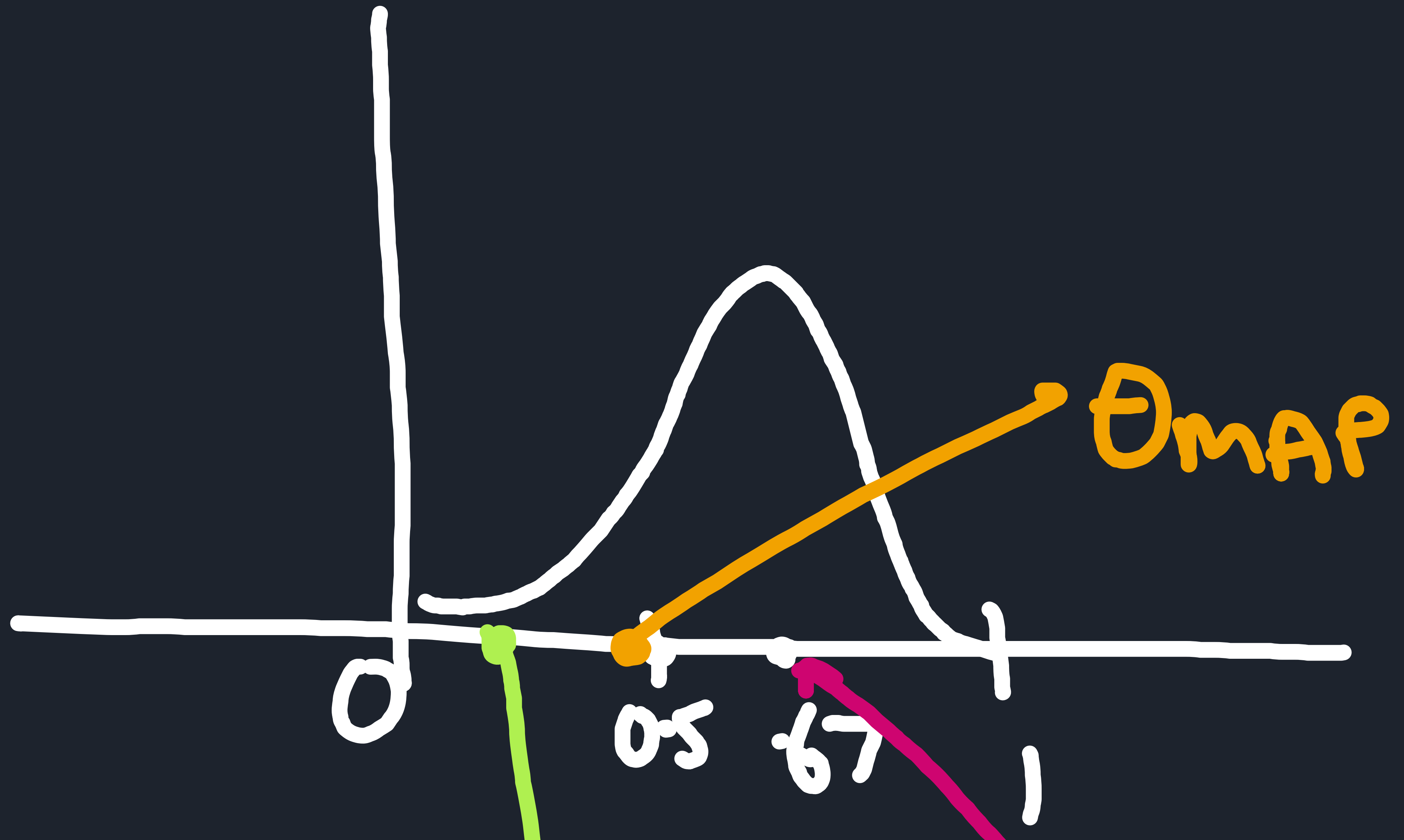
COIN TOSS MAP

$\theta = P(H)$



$$\mathrm{Beta}(\alpha,\beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

$$\theta_{\text{MODE}} = \frac{\alpha-1}{\alpha+\beta-2}$$



$\theta_{MLE} = .2$

MODE(PRIOR) $= .67$

$$\boxed{\hat{\theta}_{MAP} = \frac{n_H + \alpha - 1}{n_H + \alpha - 1 + n_T + \beta - 1}}$$

$\alpha = 11 \; ; \beta = 6$

$$\text{mode} = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{10}{15} = 2/3$$

$n_H = 2$

$n_T = 8$

$$\hat{\theta}_{MLE} = \frac{n_H}{n_H + n_T}$$