

BAYES RULE

Prior

BAYES RULE

Prior

Knowledge before |
prior to
observations

BAYES RULE

Prior

Data

Knowledge before |
prior to
observations

BAYES RULE

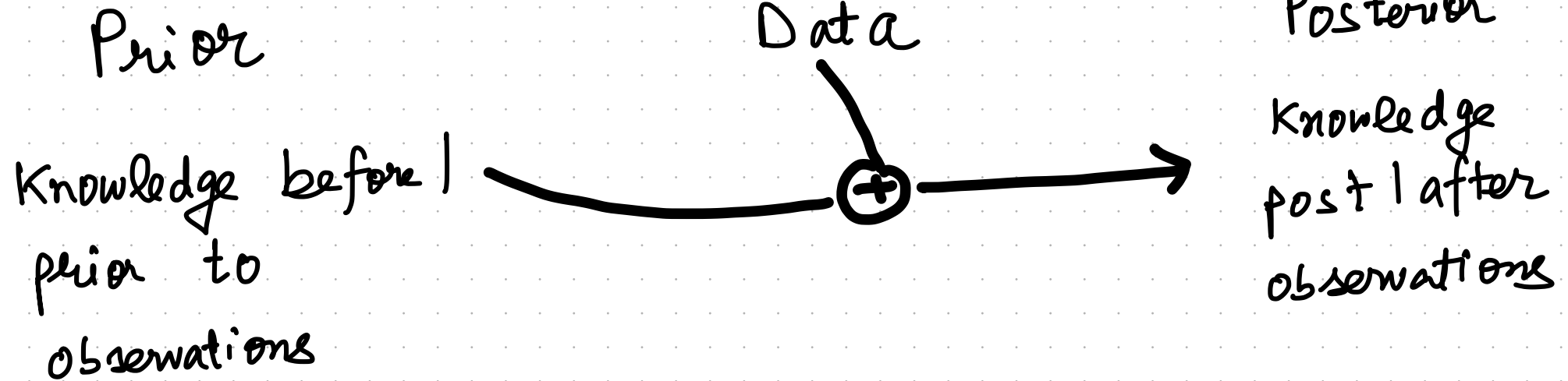
Prior

Data

Posterior

Knowledge before |
prior to
observations

BAYES RULE



BAYES RULE

Ques: Is Steve librarian or farmer?

BAYES RULE

Ques: Is Steve librarian or farmer?

Depends!
on what you use to decide?

BAYES RULE

Ques: Is Steve librarian or farmer?

Depends!

On what you use to decide?

PRIOR

DATA

POSTERIOR

$$\frac{\text{Librarians}}{\text{Farmers}} = \frac{1}{20}$$

$$P(L) / P(F) = 1/20$$

$$P(L) / P(F) = 1/20$$

BAYES RULE

Ques: Is Steve librarian or farmer?

Depends!

on what you use to decide?

DATA

POSTERIOR

PRIOR

$$\frac{\text{Librarians}}{\text{Farmers}} = \frac{1}{20}$$

MOST likely a
Farmer

BAYES RULE

NEW!! : Steve is meek, orderly, ...

Ques: Is Steve librarian(L) or farmer?
(F)

Depends!

on what you use to decide?

PRIOR

DATA

POSTERIOR

$$P(\text{Description} | L) = 0.4$$

$$P(\text{Description} | F) = 0.1$$

BAYES RULE

NEW!! : Steve is meek, orderly,

Ques: Is Steve librarian(L) or farmer?
(F)

Depends!

on what you use to decide?

PRIOR

DATA

$$P(\text{Description} | L) = 0.4$$

$$P(\text{Description} | F) = 0.1$$

MOST LIKELY A
LIBRARIAN

POSTERIOR

BAYES RULE

NEW!! : Steve is meek, orderly, ...

Ques: Is Steve librarian(L) or farmer?
(F)

Depends!

on what you use to decide?

PRIOR

DATA

POSTERIOR

- Account for prior
&
observations

BAYES RULE

NEW!! : Steve is meek, orderly, ...

Ques: Is Steve librarian(L) or farmer? (F)

Depends!

on what you use to decide?

PRIOR

DATA

POSTERIOR

- Account for prior & observations
- $P(L | \text{Description}) = ?$
- $P(F | \text{Description}) = ?$

BAYES RULE

NEW!! : Steve is meek, orderly, ...

Ques: Is Steve librarian(L) or farmer?
(F)

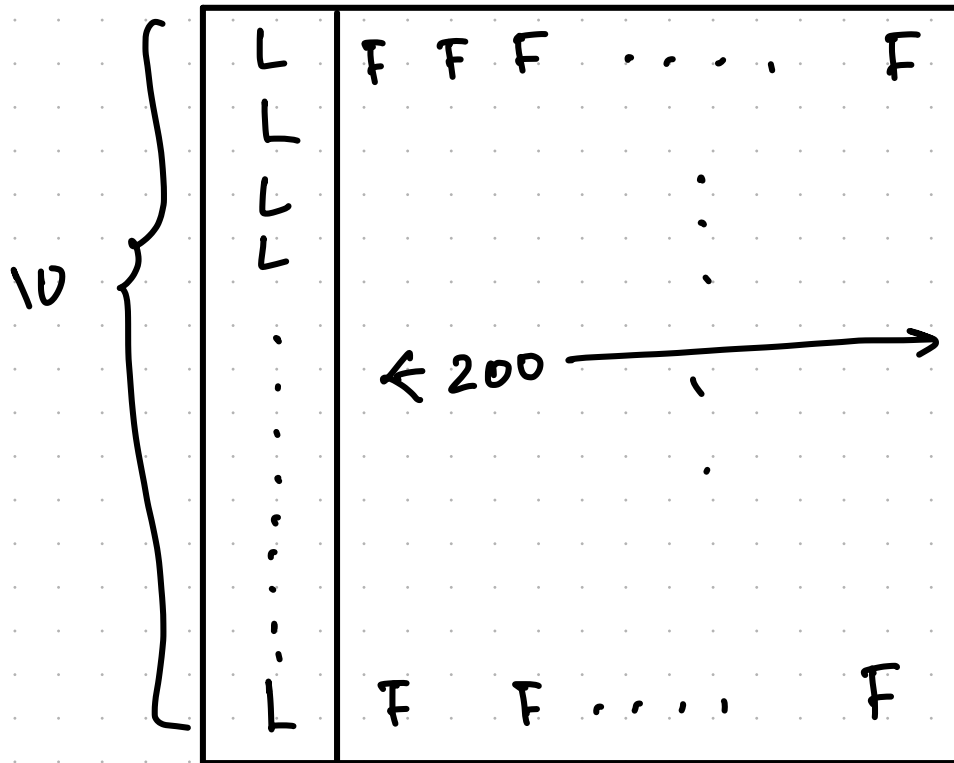
- $P(L | \text{Description}) = ?$

BAYES RULE

NEW!! : Steve is meek, orderly,

Ques: Is Steve librarian(L) or farmer?(F)

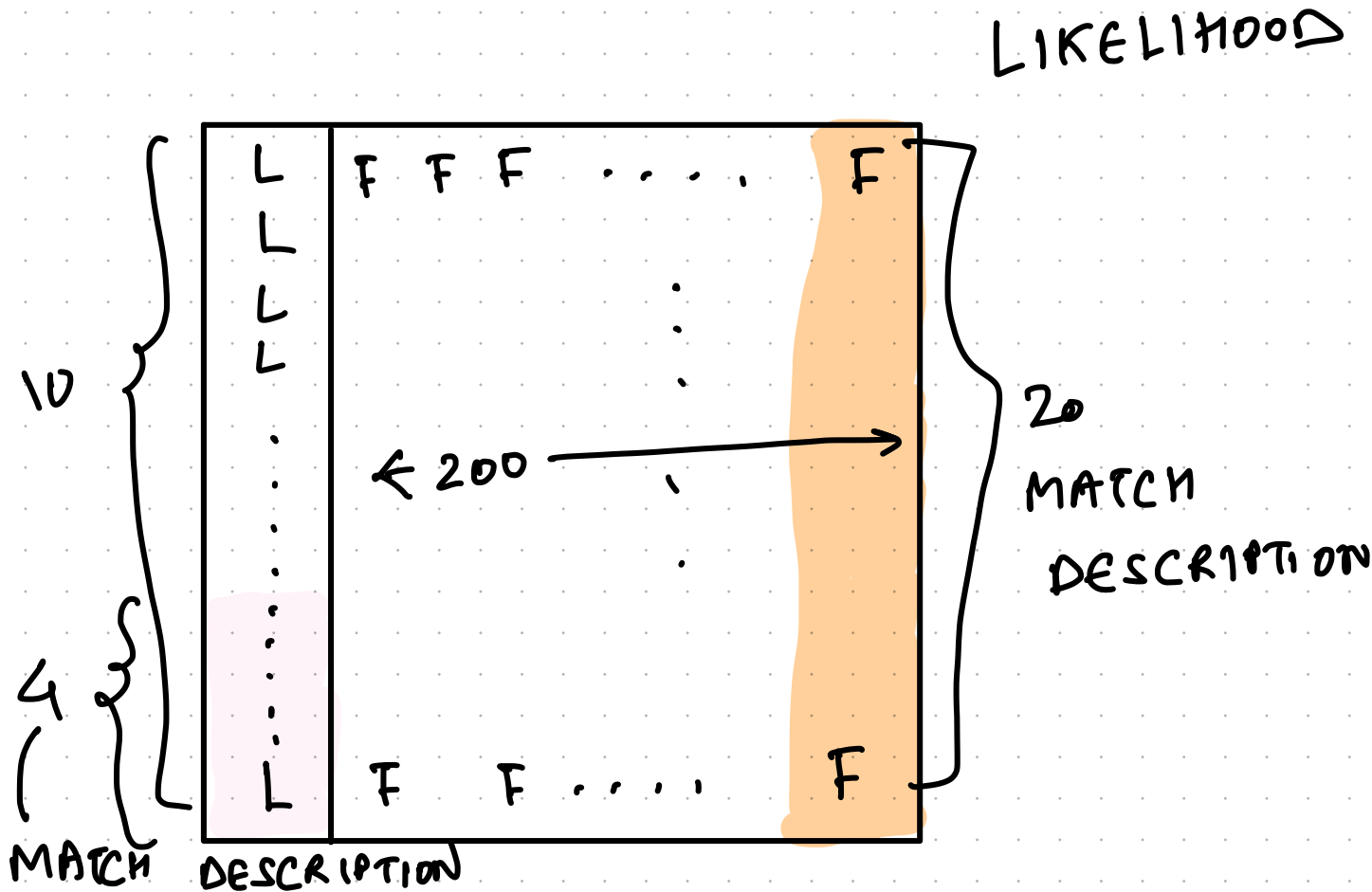
PRIOR



BAYES RULE

NEW!! : Steve is meek, orderly, ...

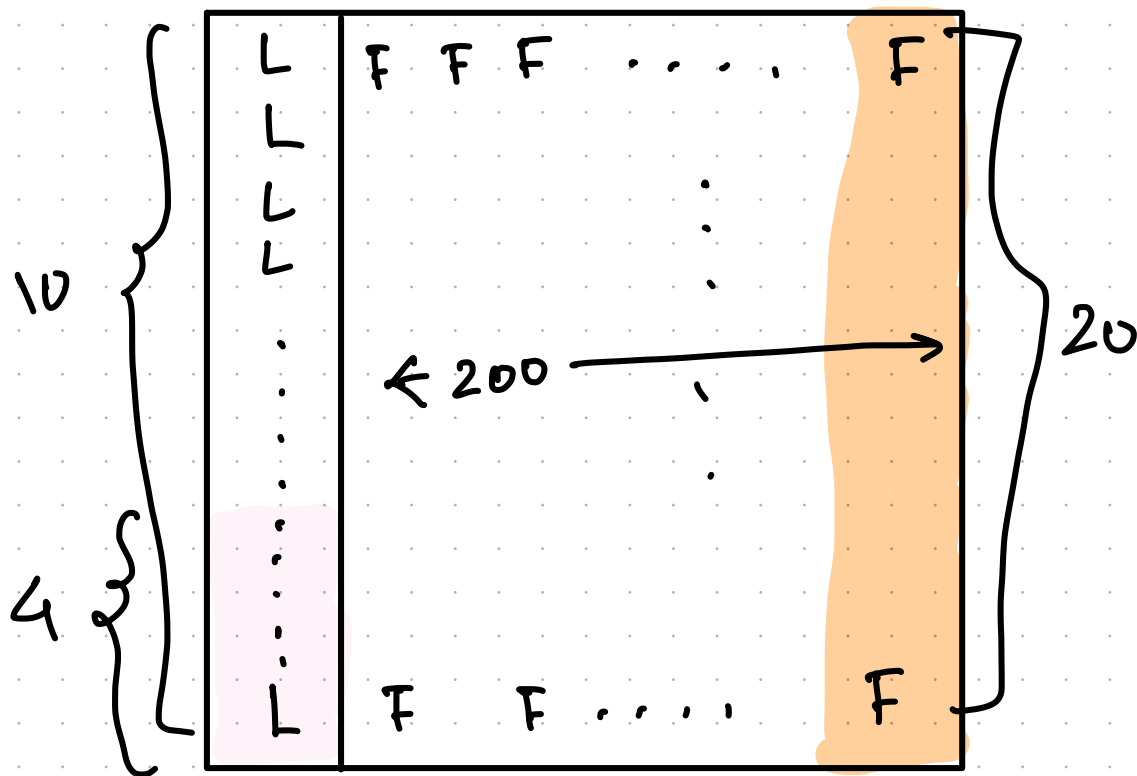
Ques: Is Steve librarian(L) or farmer? (F)



BAYES RULE

NEW!! : Stone is meek, orderly, ...

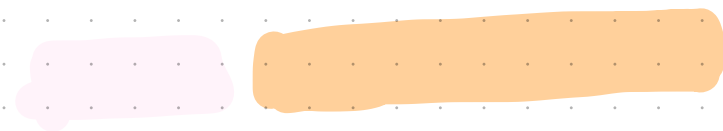
Ques: Is Stone liberarian (L) or farmer? (F)



POSTERIOR

$P(L | \text{Description})$

=

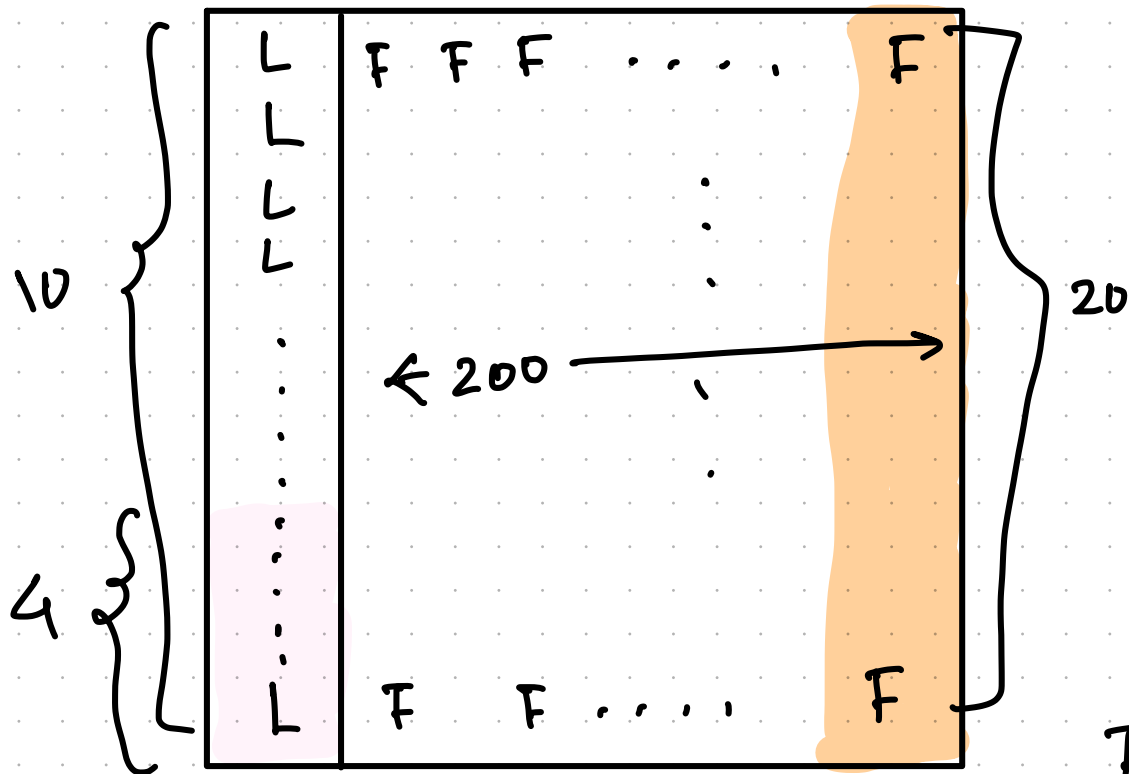


$$= \frac{4}{4+20} = \frac{1}{6}$$

BAYES RULE

NEW!! : Steve is meek, orderly, ...

Ques: Is Steve liberarian (L) or farmer? (F)



POSTERIOR

$P(L | \text{Description})$

$$= 4$$

$$4 + 20$$

$$= \frac{4}{4+20} = \frac{1}{6}$$

FARMER MORE LIKELY

BAYES RULE FOR ML

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

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

* BAYES RULE FOR ML

Bayesian Linear Regression

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

← NORMAL (pointing to $P(D | \theta)$)

← NORMAL (pointing to $P(\theta)$)

← NORMAL (pointing to $P(\theta | D)$)

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

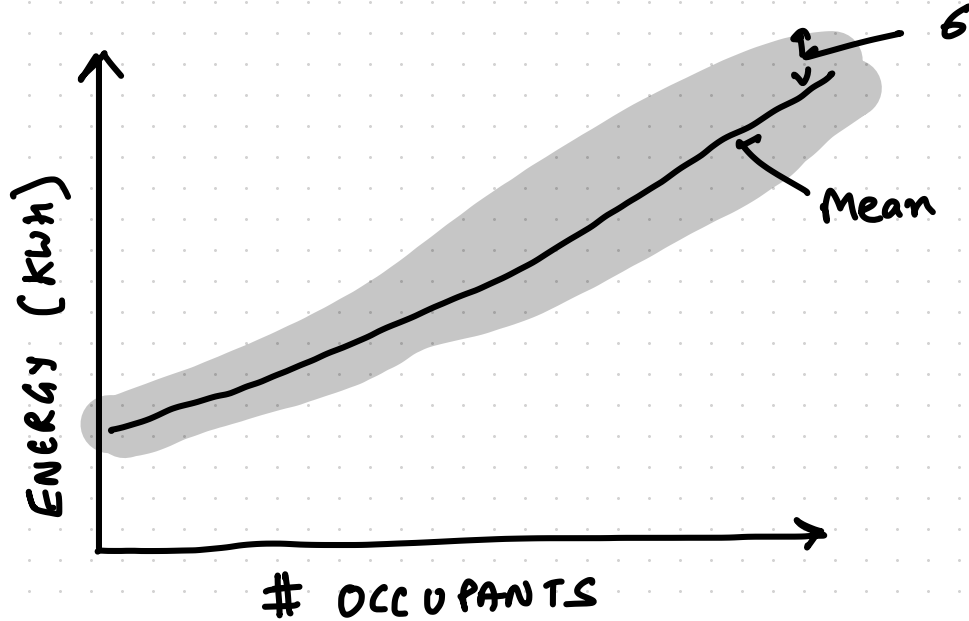
* : DETAILS LATER! JUST SHOWN AS AN EXAMPLE

IITGN new hostel electricity consumption (LINEAR MODEL)



IITGN new hostel electricity consumption (LINEAR MODEL)

GOAL



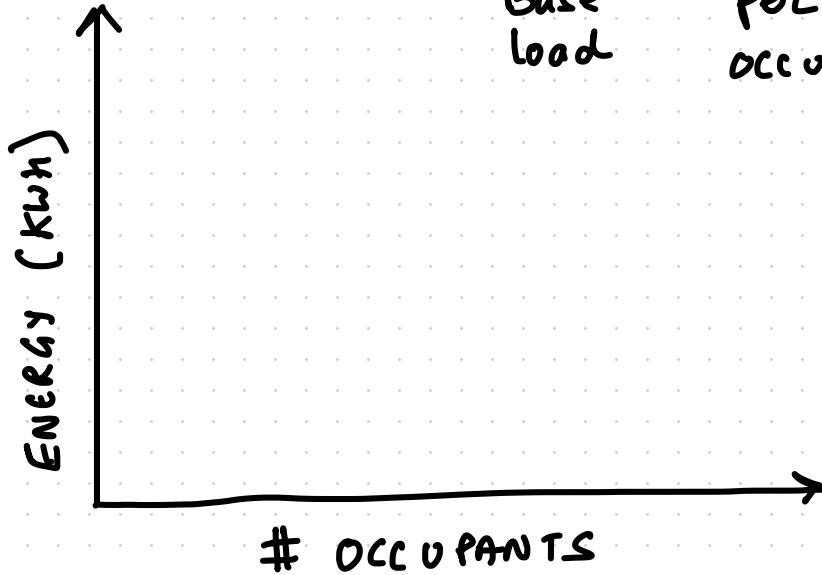
IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : No Data
- Params: θ_0, θ_1



IITGN new hostel electricity consumption (LINEAR MODEL)

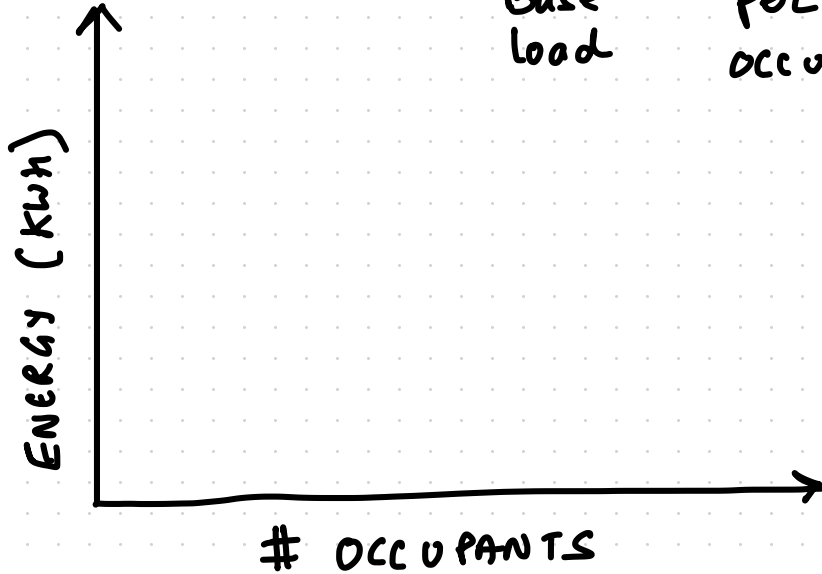
- New hostel : No Data
- Params : θ_0, θ_1
 - Base load
 - Consumption per occupant



IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : No Data
- Params : θ_0, θ_1
 - Base load
 - Consumption per occupant

PRIOR?



IITGN new hostel electricity consumption (LINEAR MODEL)

- θ_1 : +ve blw 1 & 2
- θ_0 : +ve, k units



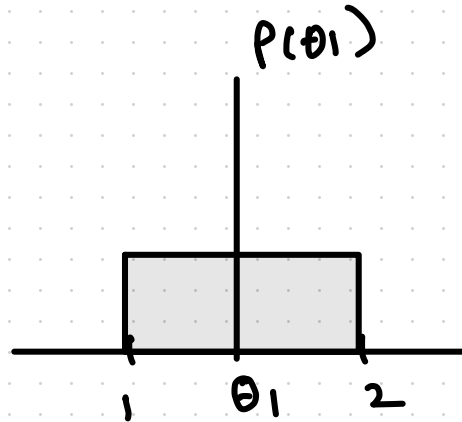
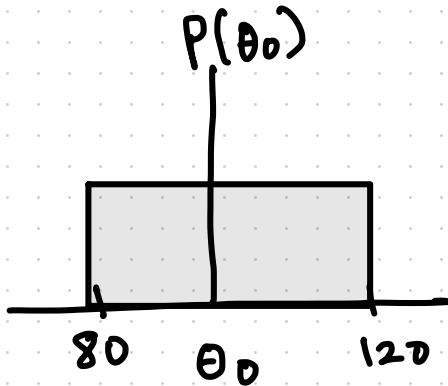
IITGN new hostel electricity consumption (LINEAR MODEL)

- θ_1 : +ve blw 1 & 2
- θ_0 : +ve, K units (say 100)

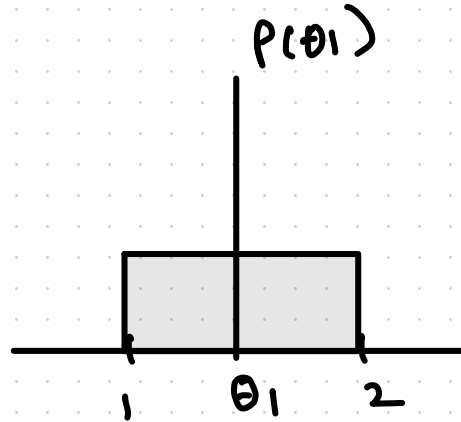
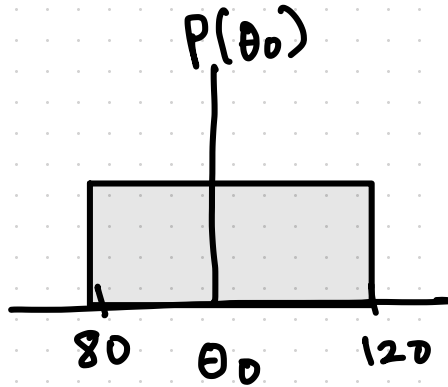
IITGN new hostel electricity consumption (LINEAR MODEL)

- θ_1 : tve blw 1 & 2
- θ_0 : tve, K units (say 100)

UNIFORM PRIOR

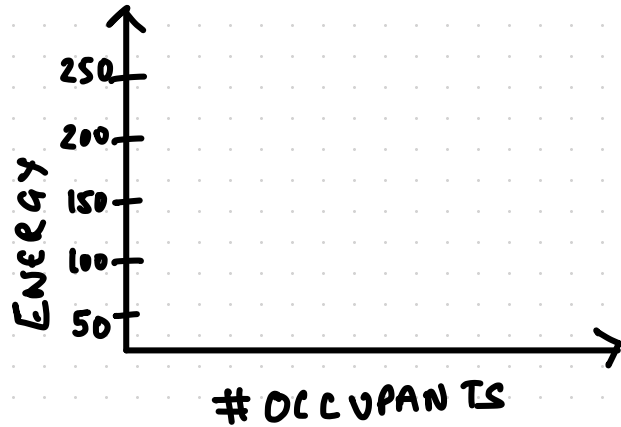
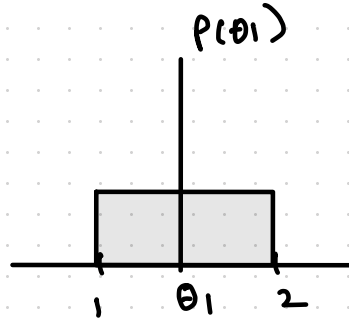
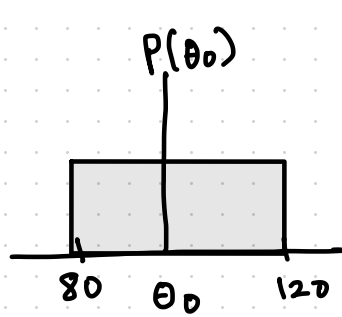


IITGN new hostel electricity consumption (LINEAR MODEL)
UNIFORM PRIOR SAMPLES



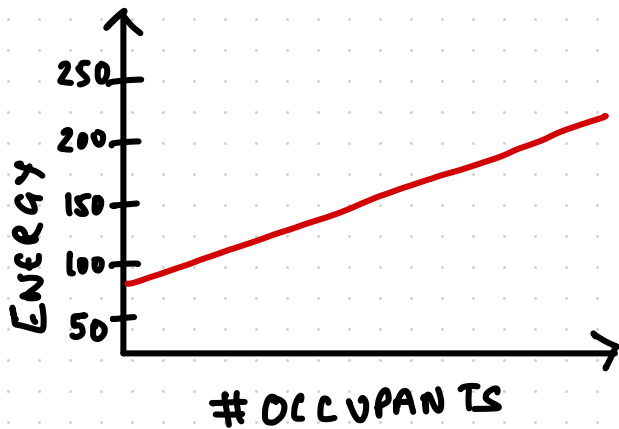
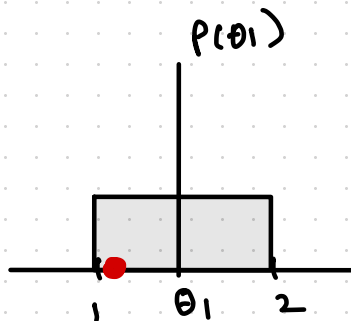
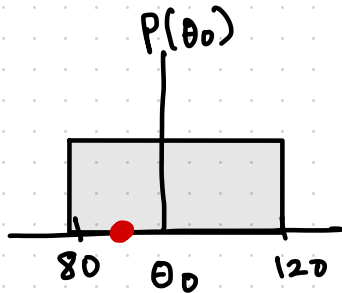
IITGN new hostel electricity consumption (LINEAR MODEL)

UNIFORM PRIOR SAMPLES



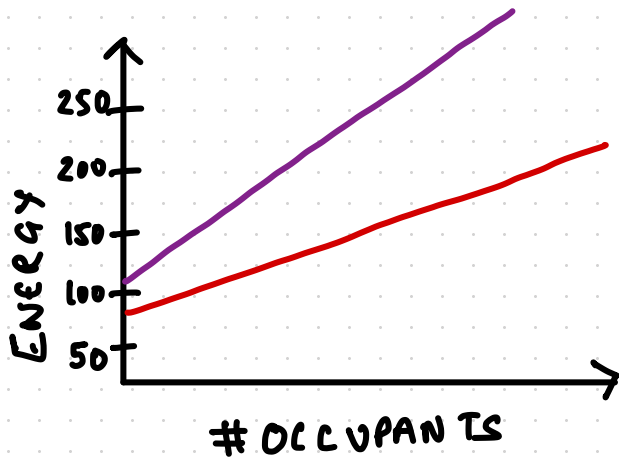
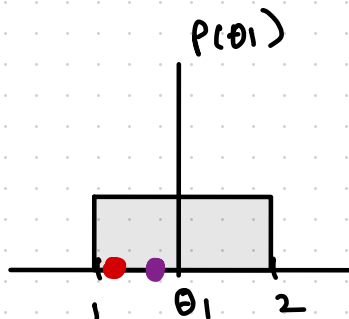
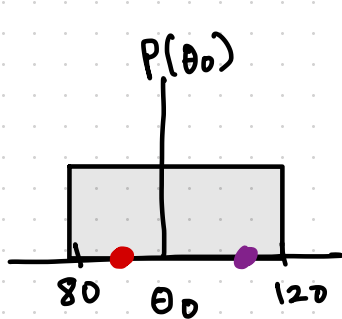
IITGN new hostel electricity consumption (LINEAR MODEL)

UNIFORM PRIOR SAMPLES



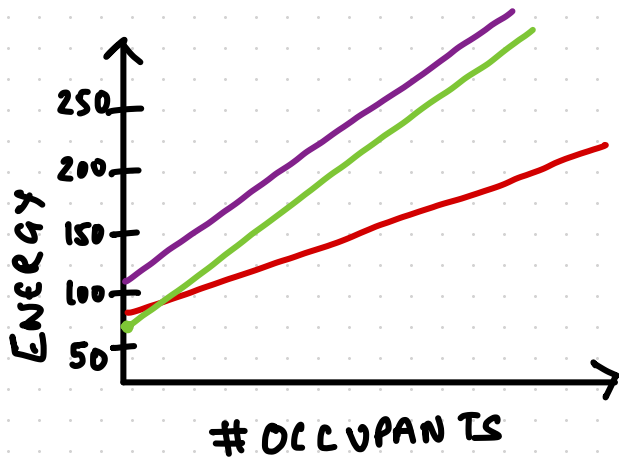
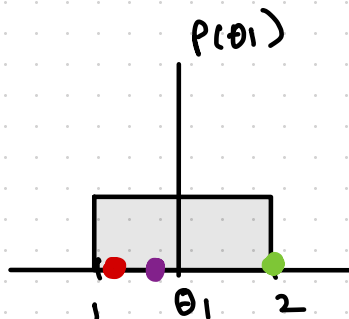
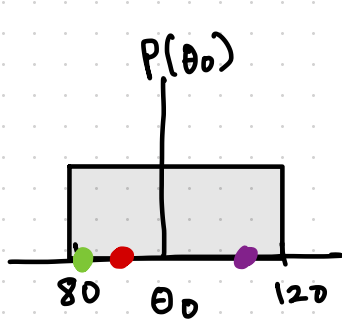
IITGN new hostel electricity consumption (LINEAR MODEL)

UNIFORM PRIOR SAMPLES



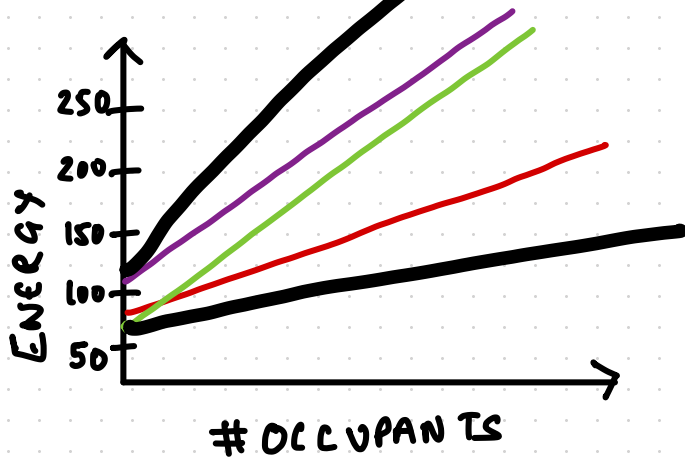
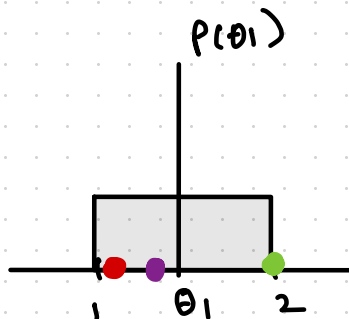
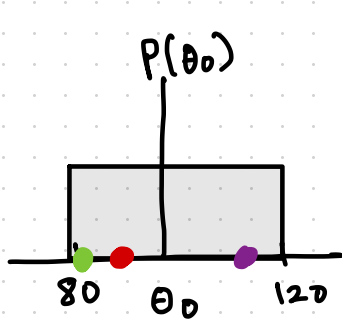
IITGN new hostel electricity consumption (LINEAR MODEL)

UNIFORM PRIOR SAMPLES



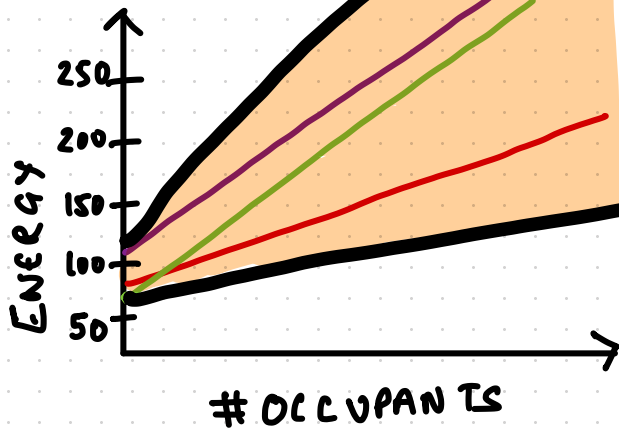
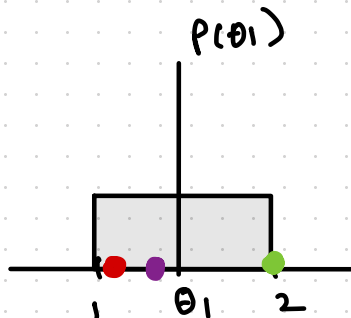
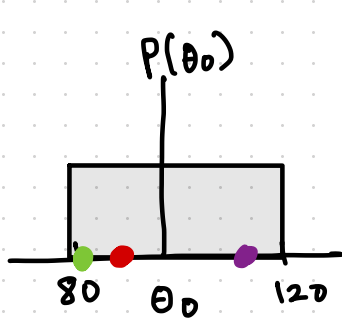
IITGN new hostel electricity consumption (LINEAR MODEL)

UNIFORM PRIOR SAMPLES



IITGN new hostel electricity consumption (LINEAR MODEL)

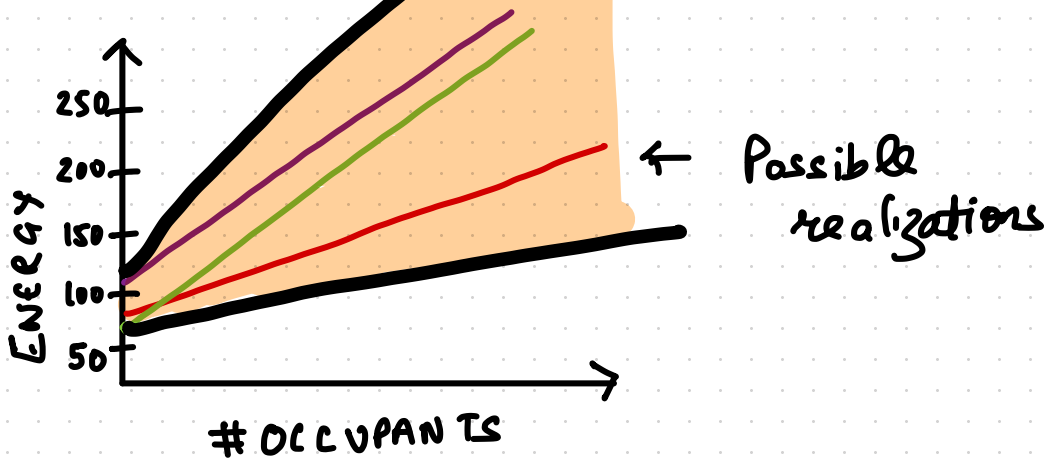
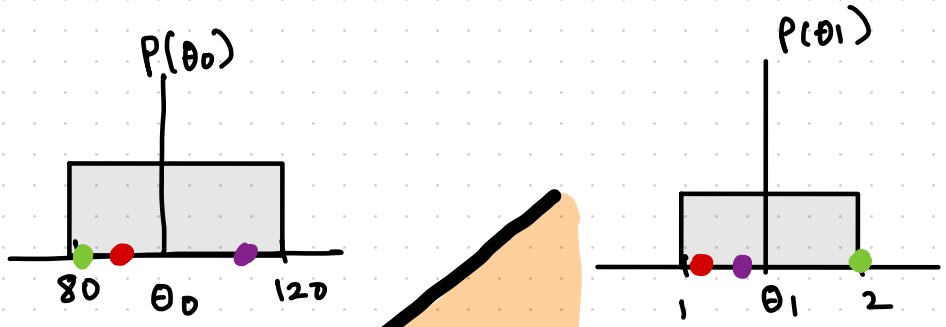
UNIFORM PRIOR SAMPLES



← Possible realizations

IITGN new hostel electricity consumption (LINEAR MODEL)

UNIFORM PRIOR SAMPLES



NOTEBOOK

MONTE CARLO ESTIMATION

- CONGRATS! you implemented M.C. estimation already!

MONTE CARLO ESTIMATION

- CONGRATS! you implemented M.C. estimation already?

- We wanted

$$p(y^* | x^*) = ?$$

MONTE CARLO ESTIMATION

- CONGRATS! you implemented M.C. estimation already!

- We wanted

$$P(y^* | x^*) = ?$$

$$= \int P(y^*, \theta | x^*) d\theta \quad [\text{Marginalisation}]$$

MONTE CARLO ESTIMATION

- CONGRATS! you implemented M.C. estimation already!

- We wanted

$$P(y^* | x^*) = ?$$

$$= \int P(y^*, \theta | x^*) d\theta$$

[Marginalisation]

$$= \int P(y^* | x^*, \theta) \cdot P(\theta) d\theta$$

[Bayes rule]

MONTE CARLO ESTIMATION

- CONGRATS! you implemented M.C. estimation already?

- We wanted

$$P(y^* | x^*) = ?$$

$$= \int P(y^*, \theta | x^*) d\theta$$

[Marginalisation]

$$= \int P(y^* | x^*, \theta) \cdot p(\theta) d\theta$$

[Bayes rule]

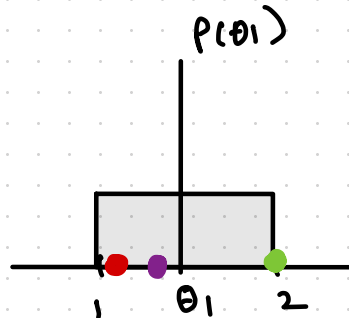
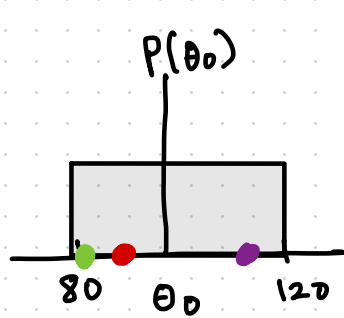
$$= \sum_{s=1}^S \frac{P(y^* | x^*, \theta_s)}{S}$$

[Monte Carlo]

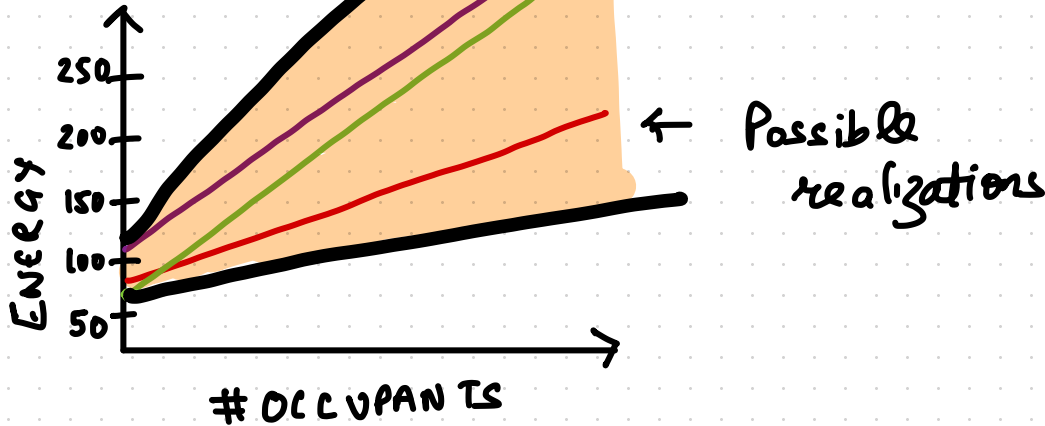
where $\theta_s \sim p(\theta)$

IITGN new hostel electricity consumption (LINEAR MODEL)

UNIFORM PRIOR SAMPLES

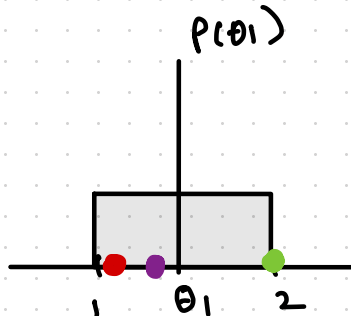
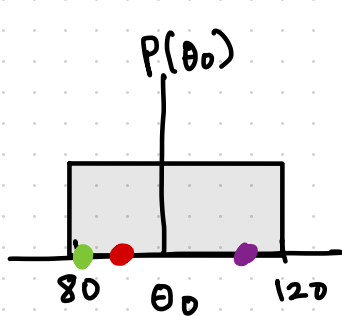


LIMITATION OF USING THIS PRIOR?

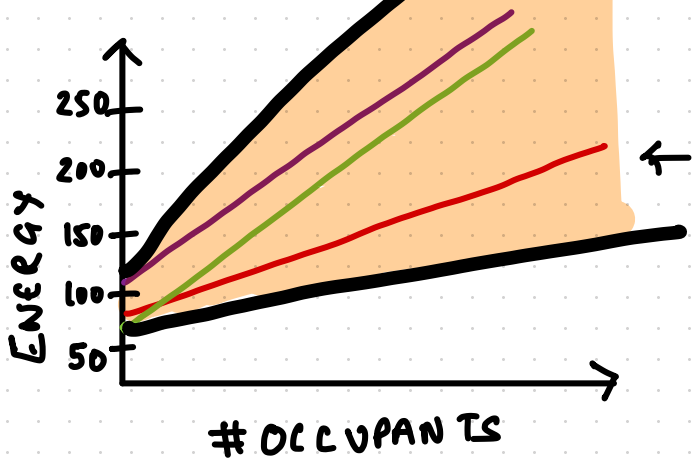


IITGN new hostel electricity consumption (LINEAR MODEL)

UNIFORM PRIOR SAMPLES



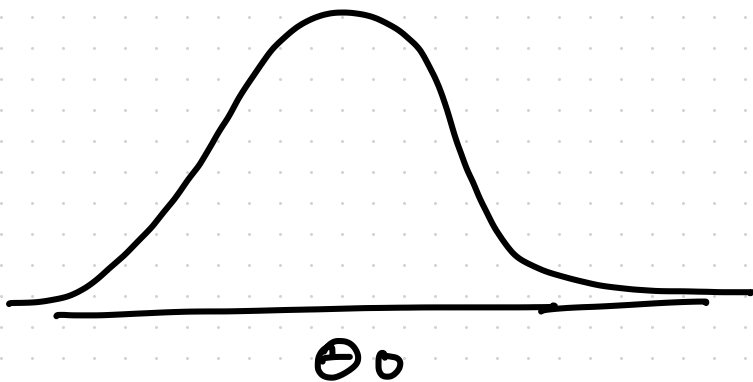
LIMITATION OF USING THIS PRIOR?



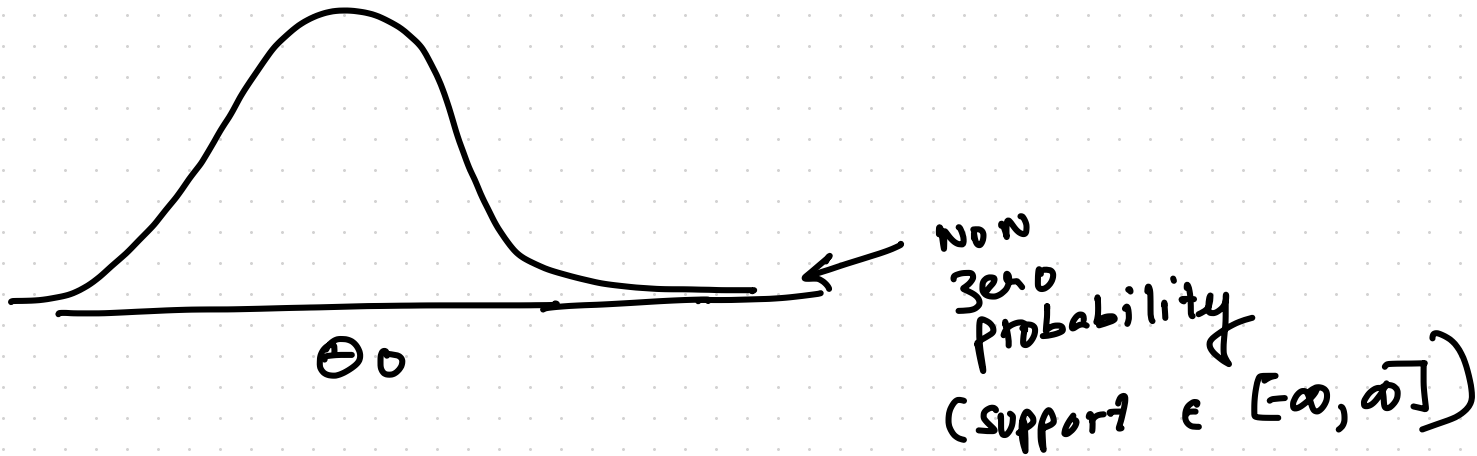
Possible realizations

$\theta_1 > 2$ not possible

IITGN new hostel electricity consumption (LINEAR MODEL)
NORMAL PRIOR SAMPLES

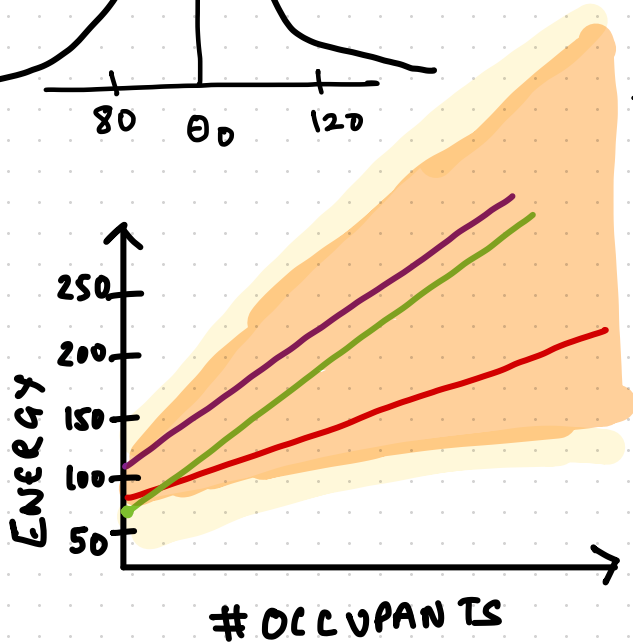
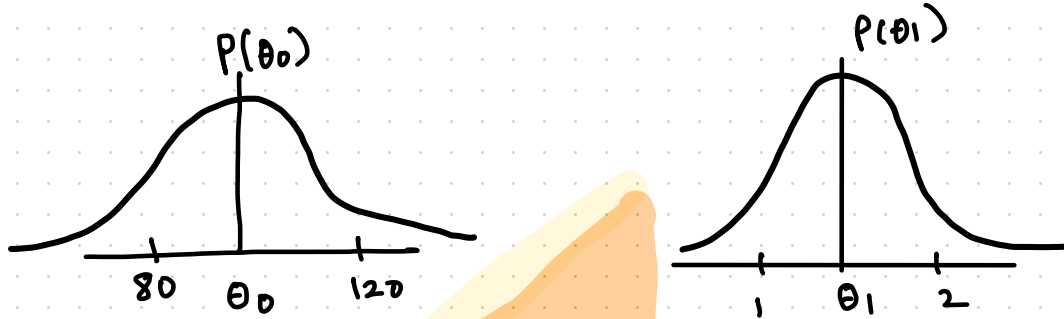


IITGN new hostel electricity consumption
(LINEAR MODEL)
NORMAL PRIOR SAMPLES



IITGN new hostel electricity consumption (LINEAR MODEL)

NORMAL PRIOR SAMPLES



IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : No Data
- Params: θ_0, θ_1



BEST Estimate
for
 θ_0, θ_1
w/o
DATA?

IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : No Data
- Params: θ_0, θ_1

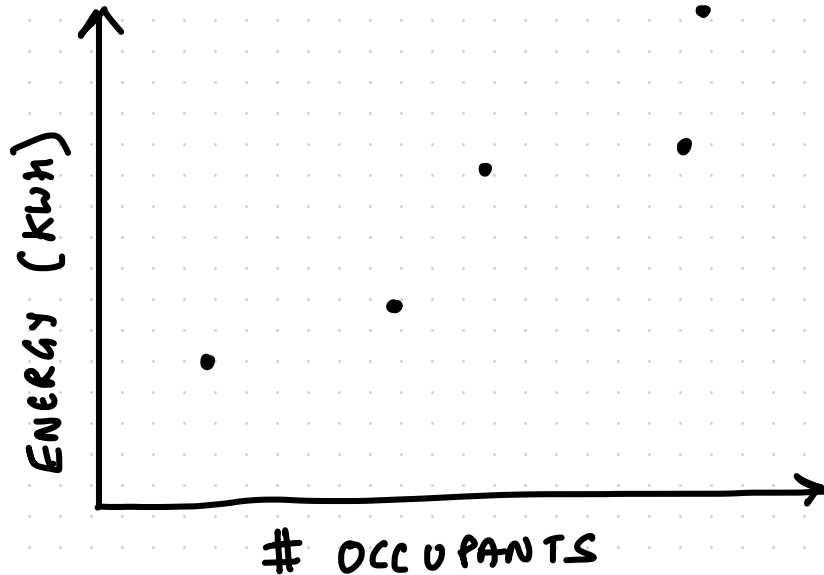


BEST estimate
for
 θ_0, θ_1
w/o
DATA?

PRIOR
eg. $\theta_0 \sim N(100, 1)$

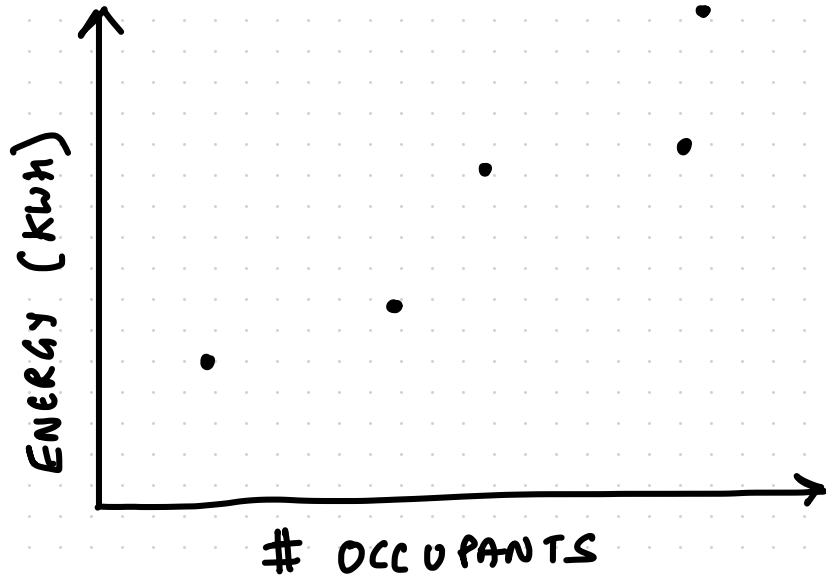
IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : ✓ Data
- Params: θ_0, θ_1



IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : ✓ Data
- Params: θ_0, θ_1

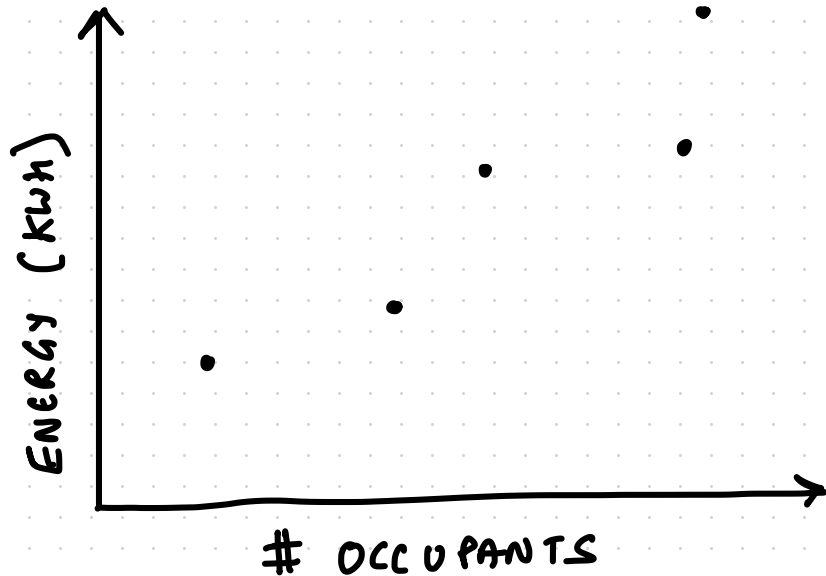


- For a moment forget prior
- What is best guess / estimate for (θ_0, θ_1) ?

IITGN new hostel electricity consumption (LINEAR MODEL)

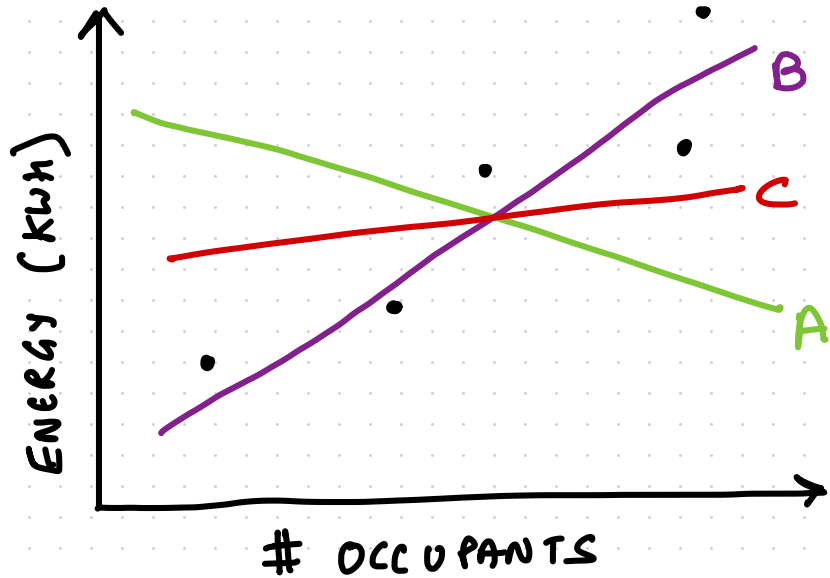
- New hostel : ✓ Data
- Params: θ_0, θ_1

- what is best guess / estimate for (θ_0, θ_1) ?



IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : ✓ Data
- Params: θ_0, θ_1

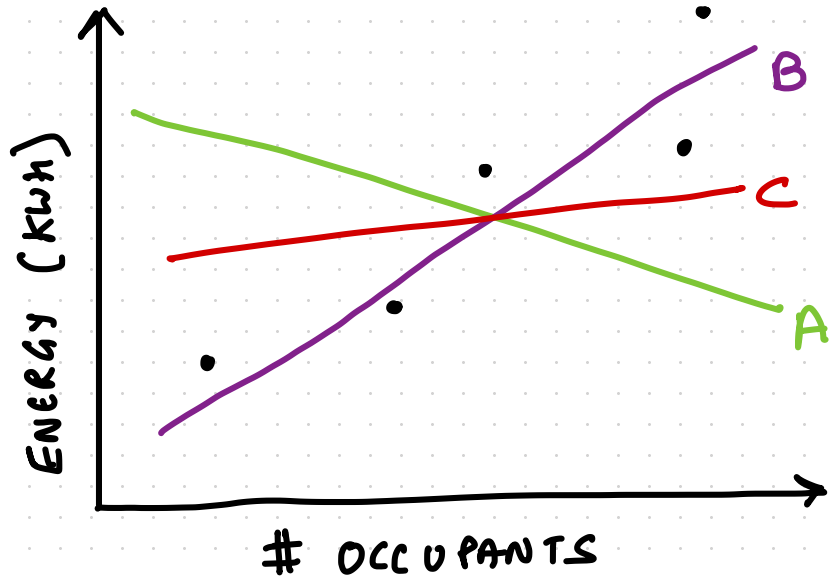


- what is best guess
| estimate for (θ_0, θ_1) ?

- A or B or C?

IITGN new hostel electricity consumption (LINEAR MODEL)

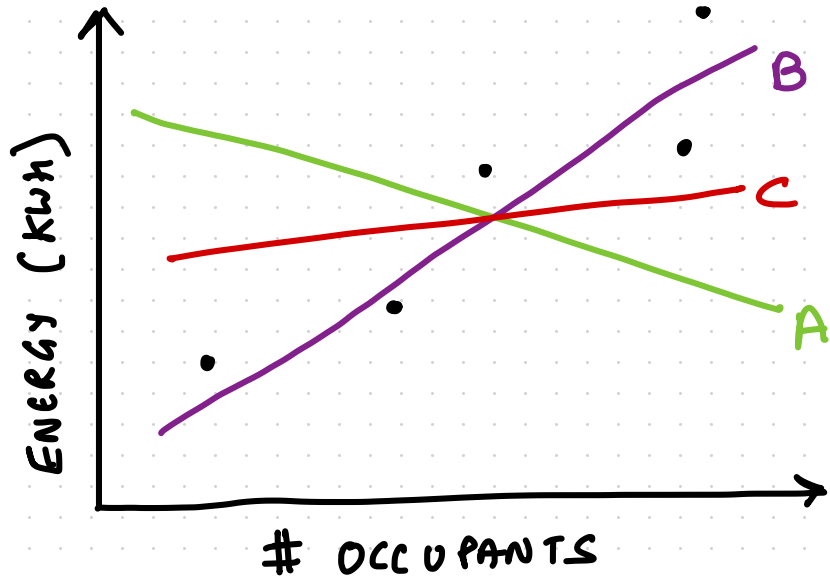
- New hostel : ✓ Data
- Params: θ_0, θ_1



- what is best guess / estimate for (θ_0, θ_1) ?
- A or B or C?
- B!
- Best data "explains"

IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : ✓ Data
- Params: θ_0, θ_1



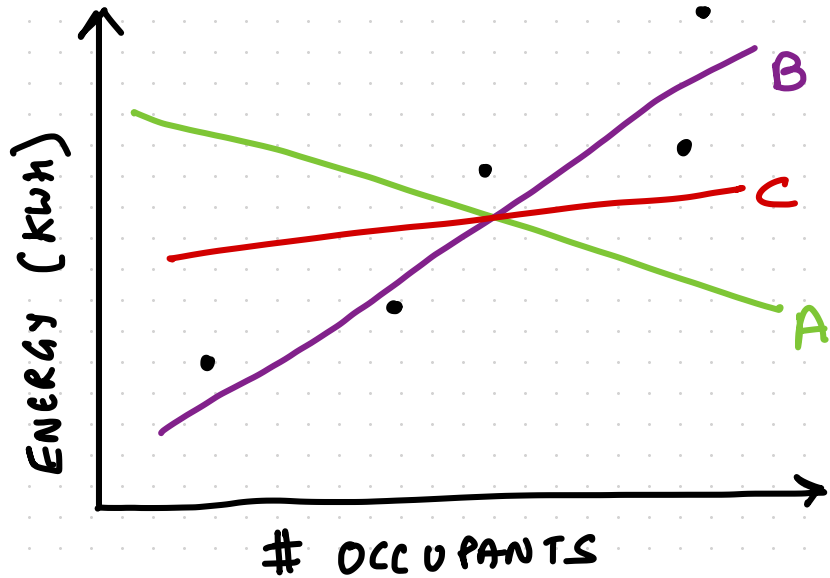
- what we want

- choose θ s.t.

$P(D|\theta)$ is max.
(Likelihood)

IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : ✓ Data
- Params: θ_0, θ_1



- what we want

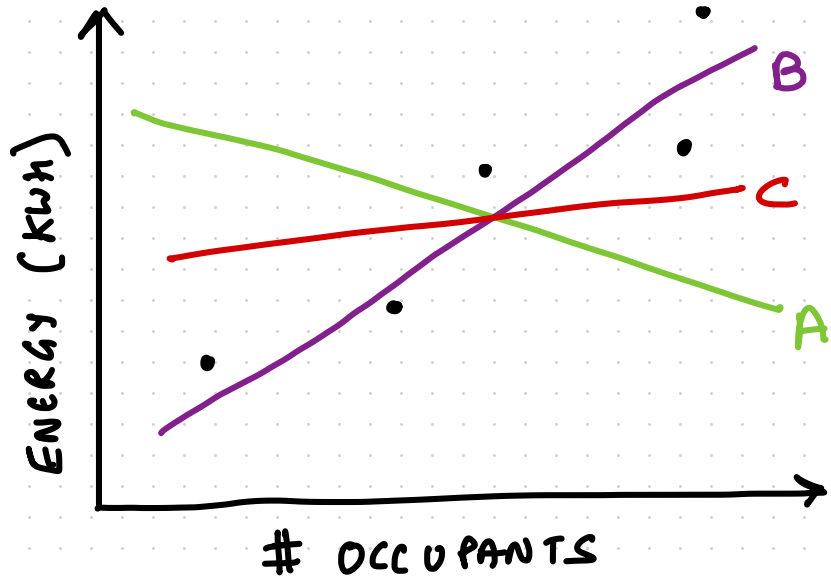
- choose θ s.t.

$P(D|\theta)$ is max.
(Likelihood)

$$\theta_{MLE} = \underset{\theta}{\operatorname{argmax}} P(D|\theta)$$

IITGN new hostel electricity consumption (LINEAR MODEL)

- New hostel : ✓ Data
- Params: θ_0, θ_1



$$\begin{aligned} - \Theta_{MLE} &= \underset{\theta}{\operatorname{argmin}} P(D|\theta) \\ &= \underset{\theta}{\operatorname{argmin}} -P(D|\theta) \\ &= \underset{\theta}{\operatorname{argmin}} -\log P(D|\theta) \\ &= \underset{\theta}{\operatorname{argmin}} L(\theta) \end{aligned}$$

IITGN new hostel electricity consumption (LINEAR MODEL)

BEST estimate
for
 θ_0, θ_1
w/o
DATA?

PRIOR
eg. $\theta_0 \sim N(100, 1)$

Best estimate
for
 θ_0, θ_1
ONLY considering
data

θ_{MLE}

IITGN new hostel electricity consumption (LINEAR MODEL)

BEST estimate
for
 θ_0, θ_1
w/o
DATA?

PRIOR
eg. $\theta_0 \sim N(100, 1)$

Best estimate
for
 θ_0, θ_1
ONLY considering
data

θ_{MLE}

IITGN new hostel electricity consumption (LINEAR MODEL)

BEST estimate
for
 θ_0, θ_1
w/o
DATA?

PRIOR
eg. $\theta_0 \sim N(100, 1)$

DISTRIBUTION

Best estimate
for
 θ_0, θ_1
ONLY considering
data

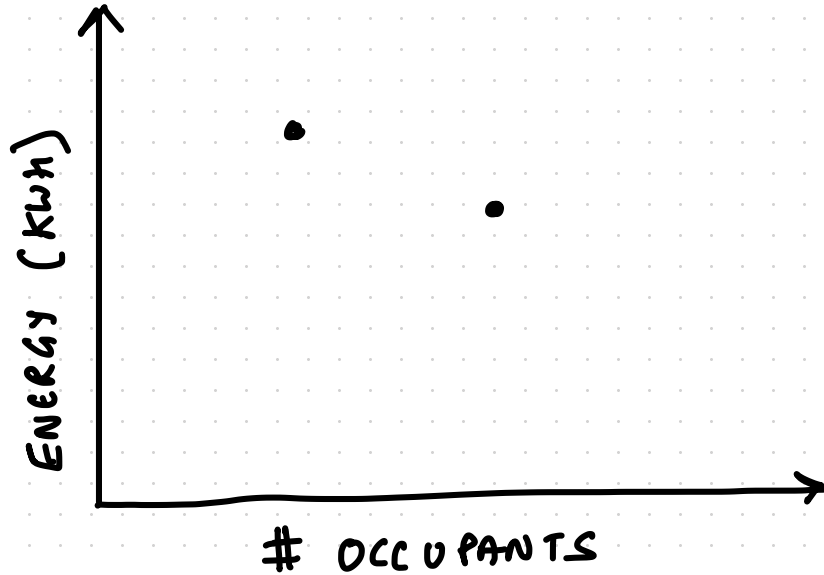
θ_{MLE}

POINT
ESTIMATE

IITGN new hostel electricity consumption (LINEAR MODEL)

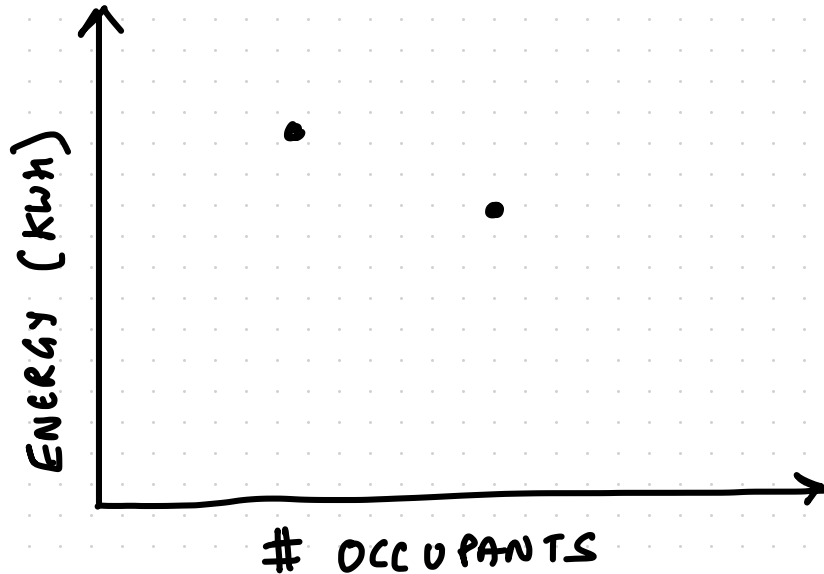
- Newer hostel : LIMITED Data
- Params: θ_0, θ_1

- what is best guess / estimate for (θ_0, θ_1) ?



IITGN new hostel electricity consumption (LINEAR MODEL)

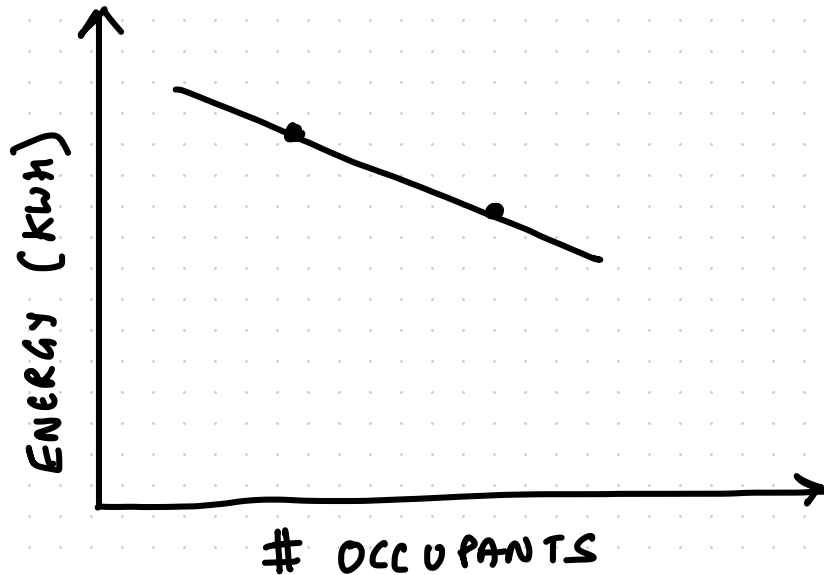
- Newer hostel : LIMITED Data
- Params: θ_0, θ_1



- what is best guess / estimate for (θ_0, θ_1) ?
- Say we consider DATA only

IITGN new hostel electricity consumption (LINEAR MODEL)

- Newer hostel : LIMITED Data
- Params: θ_0, θ_1



- what is best guess / estimate for (θ_0, θ_1) ?

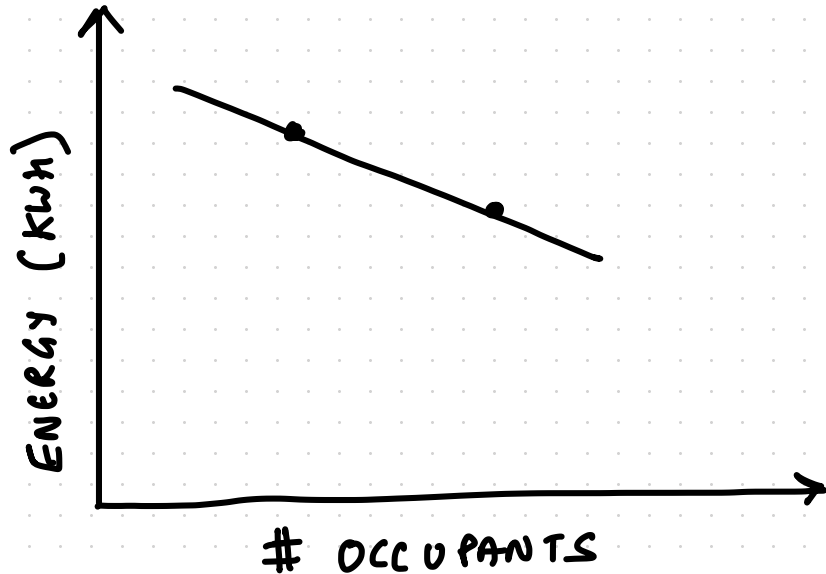
- Say we consider DATA only

- OMLE
w/
-ve slope

IITGN new hostel electricity consumption (LINEAR MODEL)

- Newer hostel : LIMITED Data
- Params: θ_0, θ_1

- what is best guess / estimate for (θ_0, θ_1) ?



- Say we consider DATA only

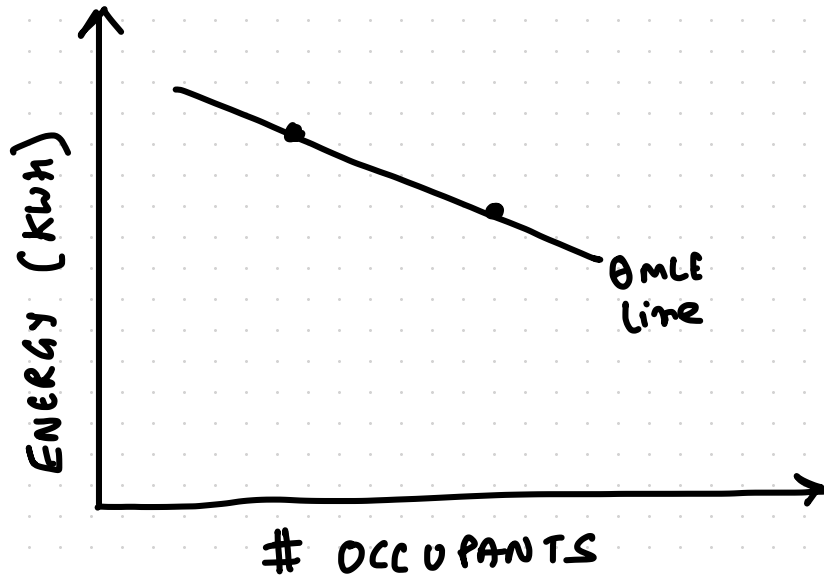
- OMLE
w/
-ve slope

- REMEDY?

IITGN new hostel electricity consumption (LINEAR MODEL)

- Newer hostel : LIMITED Data
- Params: θ_0, θ_1

- what is best guess / estimate for (θ_0, θ_1) ?



- Say we consider DATA only

- θ_{MLE}
w/
-ve slope

- REMEDY?
- CONSIDER PRIOR?

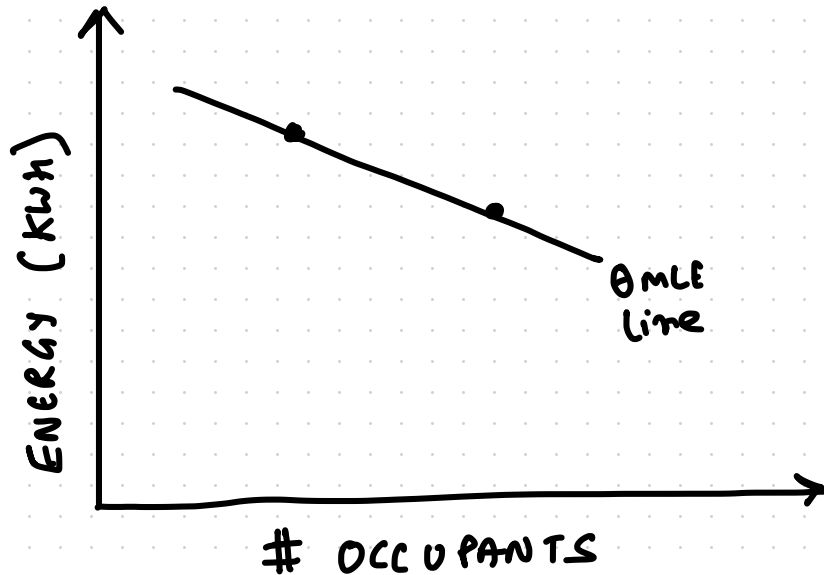
IITGN new hostel electricity consumption (LINEAR MODEL)

- Newer hostel : LIMITED Data
- Params: θ_0, θ_1

- what is best guess / estimate for (θ_0, θ_1) ?

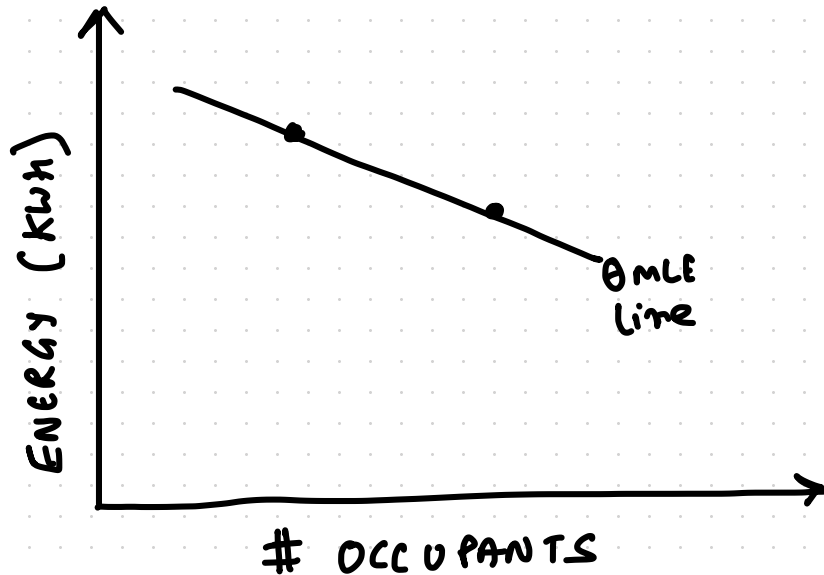
- REMEDY?
- CONSIDER PRIOR?

- $\theta_{MAP} = \underset{\theta}{\operatorname{argmax}} P(\theta | D)$



IITGN new hostel electricity consumption (LINEAR MODEL)

- Newer hostel : LIMITED Data
- Params: θ_0, θ_1



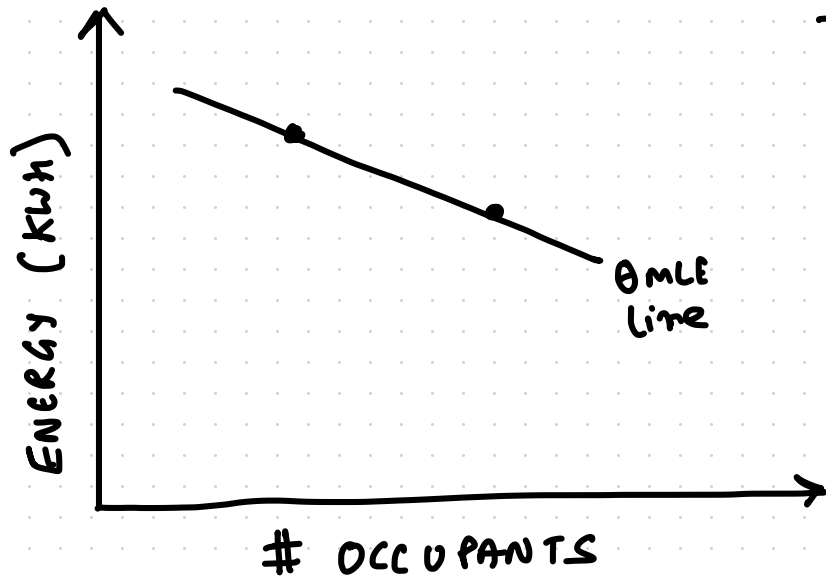
- what is best guess / estimate for (θ_0, θ_1) ?

- REMEDY?
 - CONSIDER PRIOR?

$$\begin{aligned} - \theta_{MAP} &= \underset{\theta}{\operatorname{argmax}} P(\theta | D) \\ &= \underset{\theta}{\operatorname{argmax}} \frac{P(D | \theta) \cdot P(\theta)}{\int P(D | \theta) P(\theta) d\theta} \end{aligned}$$

IITGN new hostel electricity consumption (LINEAR MODEL)

- Newer hostel : LIMITED Data
- Params: θ_0, θ_1



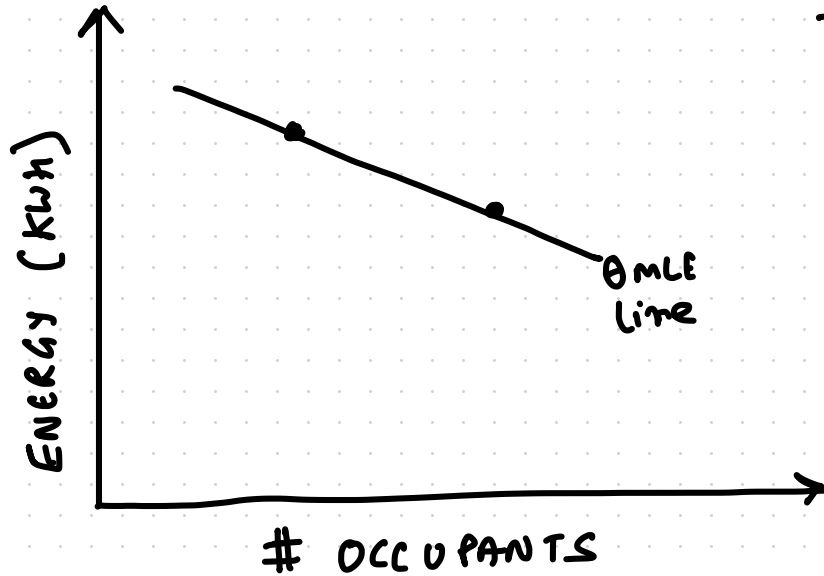
- what is best guess estimate for (θ_0, θ_1) ?

- REMEDY?
 - CONSIDER PRIOR?

$$\begin{aligned} - \theta_{\text{MAP}} &= \underset{\theta}{\text{argmax}} P(\theta | D) \\ &= \underset{\theta}{\text{argmax}} \frac{P(D | \theta) \cdot P(\theta)}{\int P(D | \theta) P(\theta) d\theta} \\ &= \underset{\theta}{\text{argmax}} P(D | \theta) \cdot P(\theta) \end{aligned}$$

IITGN new hostel electricity consumption (LINEAR MODEL)

- Newer hostel : LIMITED Data
- Params: θ_0, θ_1



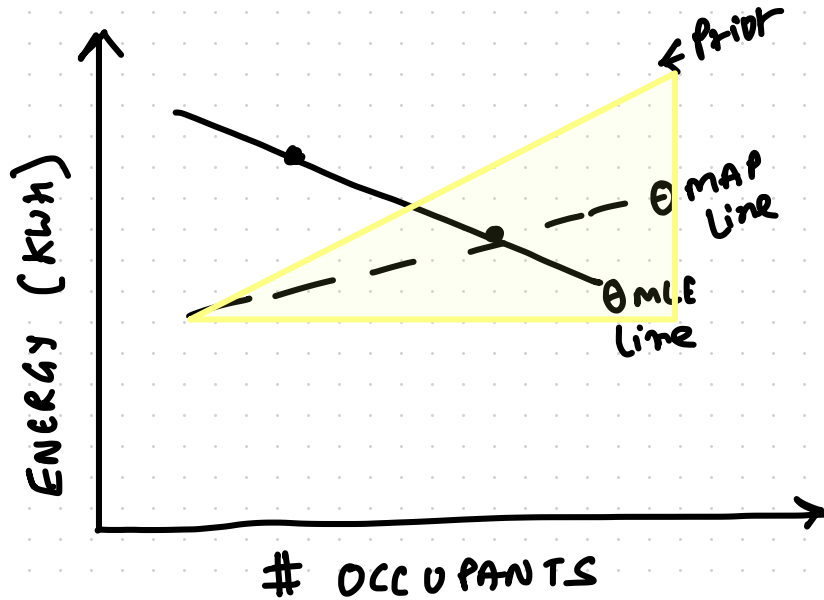
- what is best guess estimate for (θ_0, θ_1) ?

- REMEDY?
 - CONSIDER PRIOR?

$$\begin{aligned} - \theta_{\text{MAP}} &= \underset{\theta}{\text{argmax}} P(\theta | D) \\ &= \underset{\theta}{\text{argmax}} \frac{P(D|\theta) \cdot P(\theta)}{\int P(D|\theta) P(\theta) d\theta} \\ &= \underset{\theta}{\text{argmax}} \underbrace{P(D|\theta)}_{(\text{Likelihood})} \cdot \underbrace{P(\theta)}_{(\text{Prior})} \end{aligned}$$

IITGN new hostel electricity consumption (LINEAR MODEL)

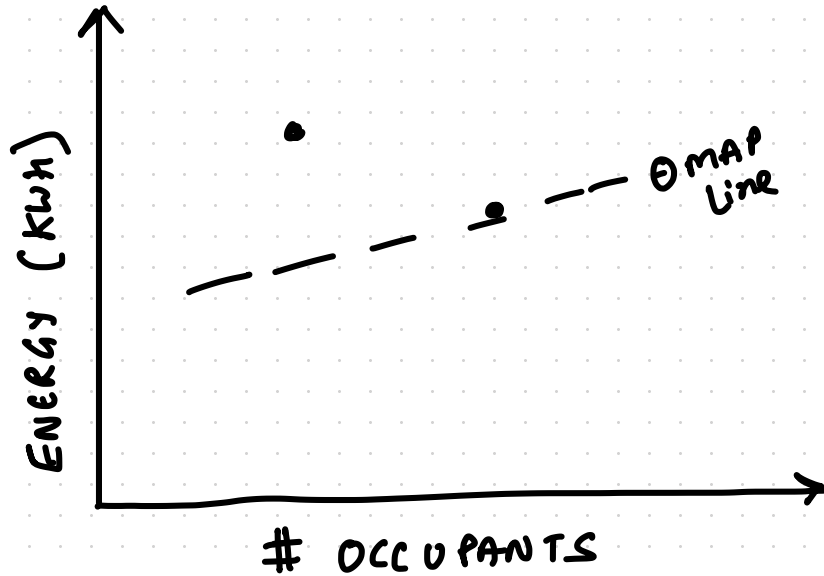
- Newer hostel : LIMITED Data
- Params: θ_0, θ_1



$$\begin{aligned} \theta_{\text{MAP}} &= \underset{\theta}{\text{argmax}} P(\theta | D) \\ &= \underset{\theta}{\text{argmax}} \frac{P(D | \theta) \cdot P(\theta)}{\int P(D | \theta) P(\theta) d\theta} \\ &= \underset{\theta}{\text{argmax}} \underbrace{P(D | \theta)}_{(\text{Likelihood})} \cdot \underbrace{P(\theta)}_{(\text{Prior})} \end{aligned}$$

IITGN new hostel electricity consumption (LINEAR MODEL)

- Newer hostel : LIMITED Data
- Params: θ_0, θ_1



$$\begin{aligned}\theta_{MAP} &= \underset{\theta}{\operatorname{argmax}} P(\theta | D) \\ &= \underset{\theta}{\operatorname{argmax}} \frac{P(D | \theta) \cdot P(\theta)}{\int P(D | \theta) P(\theta) d\theta} \\ &= \underset{\theta}{\operatorname{argmax}} \underbrace{P(D | \theta)}_{(\text{Likelihood})} \cdot \underbrace{P(\theta)}_{(\text{Prior})}\end{aligned}$$

IITGN new hostel electricity consumption (LINEAR MODEL)

BEST estimate
for
 θ_0, θ_1
w/o
DATA?

PRIOR
eg. $\theta_0 \sim N(100, 1)$

Best estimate
for
 θ_0, θ_1
ONLY considering
data

θ_{MLE}

Best estimate for
 θ_0, θ_1
CONSIDERING
- DATA
- PRIOR

θ_{MAP}

IITGN new hostel electricity consumption (LINEAR MODEL)

BEST estimate
for
 θ_0, θ_1
w/o
DATA?

PRIOR
eg. $\theta_0 \sim N(100, 1)$

DISTRIBUTION

Best estimate
for
 θ_0, θ_1
ONLY considering
data

θ_{MLE}

POINT ESTIMATE

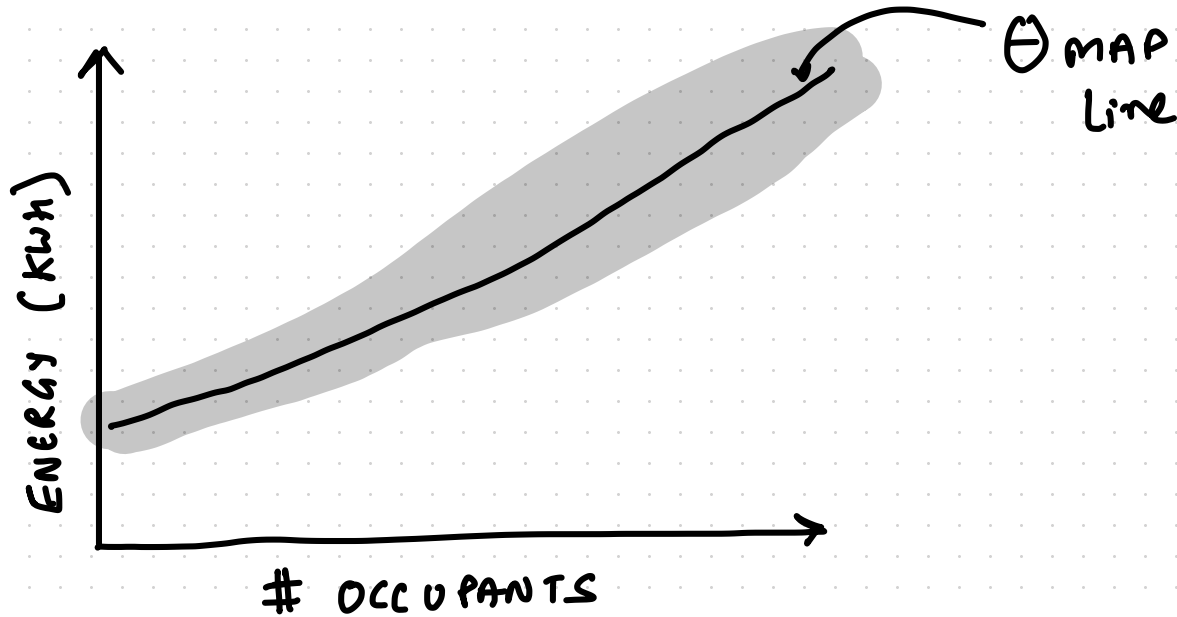
Best estimate for
 θ_0, θ_1
CONSIDERING
- DATA
- PRIOR

θ_{MAP}

POINT
ESTIMATE

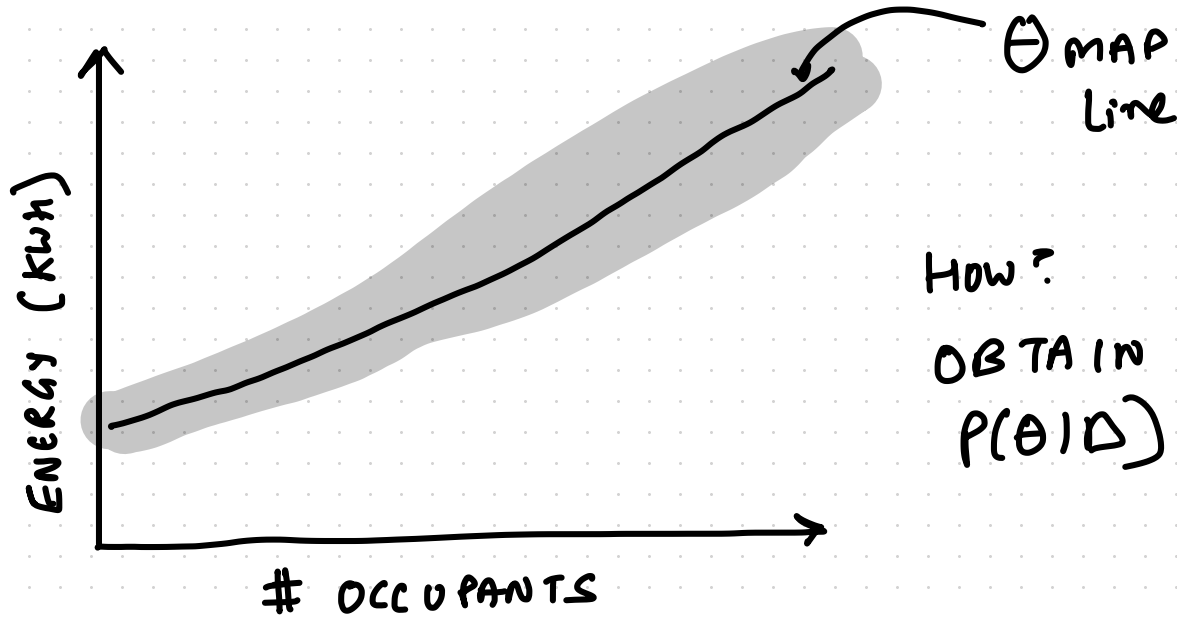
IITGN new hostel electricity consumption (LINEAR MODEL)

GOAL



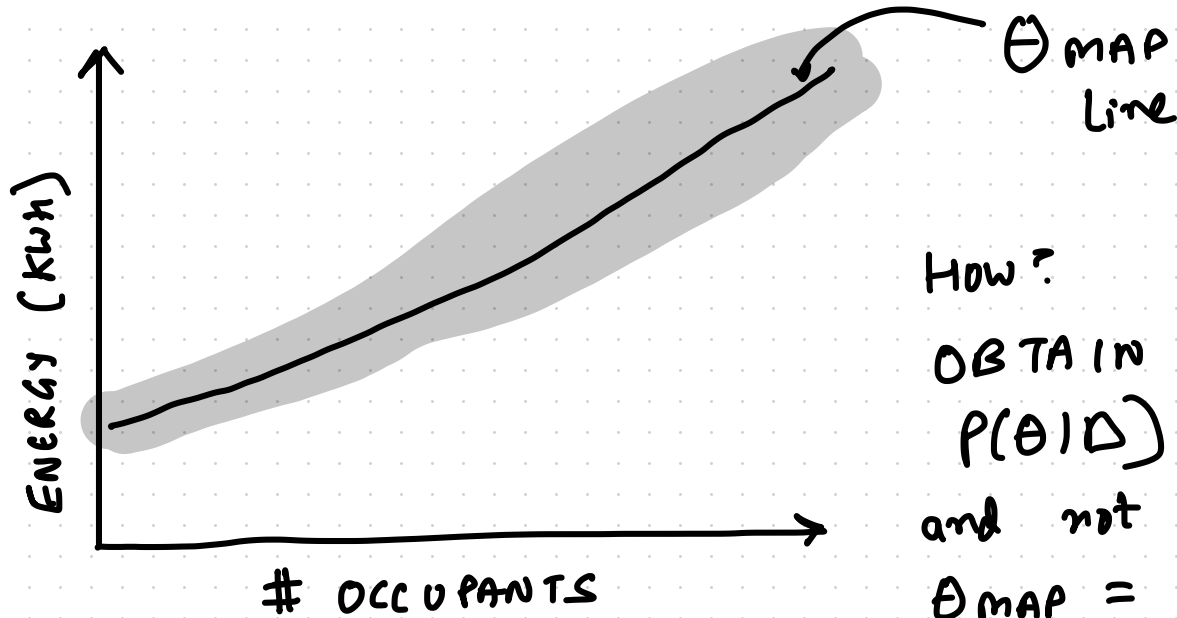
IITGN new hostel electricity consumption (LINEAR MODEL)

GOAL



IITGN new hostel electricity consumption (LINEAR MODEL)

GOAL



How?

OBTAIN
 $P(\theta | D)$

and not

$\theta_{MAP} = \underset{\theta}{\operatorname{argmax}} P(\theta | D)$

MAIN challenge

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

MAIN challenge

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

Assume:

$P(D|\theta)$ is Bernoulli ($K\theta$)

$P(\theta)$ is Normal ($\theta|\mu, \sigma$)

Now, $\int P(D|\theta) \cdot P(\theta) d\theta$ is intractable

POSTERIOR PREDICTIVE

- Assume $p(\theta|D)$ has been obtained.

- $p(y^* | x^*, D) = ?$

POSTERIOR PREDICTIVE

- Assume $p(\theta|D)$ has been obtained.

- $p(y^* | x^*, D) = ?$

$$= \int p(y^*, \theta | x^*, D) d\theta$$

[Marginalisation]

POSTERIOR PREDICTIVE

- Assume $p(\theta|D)$ has been obtained.

- $p(y^* | x^*, D) = ?$

$$= \int p(y^*, \theta | x^*, D) d\theta$$

[Marginalisation]

$$= \int p(y^* | \theta, x^*, D) \cdot p(\theta | x^*, D) d\theta$$

[Bayes Rule]

POSTERIOR PREDICTIVE

- Assume $p(\theta|D)$ has been obtained.

- $p(y^* | x^*, D) = ?$

$$= \int p(y^*, \theta | x^*, D) d\theta$$

[Marginalisation]

$$= \int p(y^* | \theta, x^*, D) \cdot p(\theta | x^*, D) d\theta$$

[Bayes Rule]

$$= \int p(y^* | \theta, x^*) \cdot p(\theta | D) d\theta$$

[Independencies]

$$= \frac{\sum_{s=1}^S p(y^* | \theta_s, x^*)}{S} \text{ where } \theta_s \sim p(\theta | D)$$

[Monte Carlo]

