Guess the weight of the fruit cake

Eric Yang

advisor:

Phil Marshall

Data from 2015 Astro Hack Week

- Phil's height, Phil's age, Phil's obesity, own age, own height, own weight
- Goal: blind inference of Phil's weight

	Timestamp	What do you think Phil's height is?	Is Phil...	What do you think Phil's age is?	What is your first name?	What is your own height?	What is your own weight?	What is your own age?	Which is the best text editor?
0	9/28/2015 17:32:57	185.0	Just right	40	John	172	148	35	Sublime
1	9/28/2015 17:47:11	195.6	Just right	44	Lia	160	130	31	emacs
2	9/28/2015 17:52:27	185.0	Just right	37	NaN	182	178	28	vim
3	9/28/2015 17:57:23	180.5	Just right	39	Rebecca	161	145	26	SUBLIME OBVIOUSLY
4	9/28/2015 18:02:57	180.0	Just right	40	Somebody	173	220	39	vim

First Method: Using crowds' estimates of Phil's height and obesity

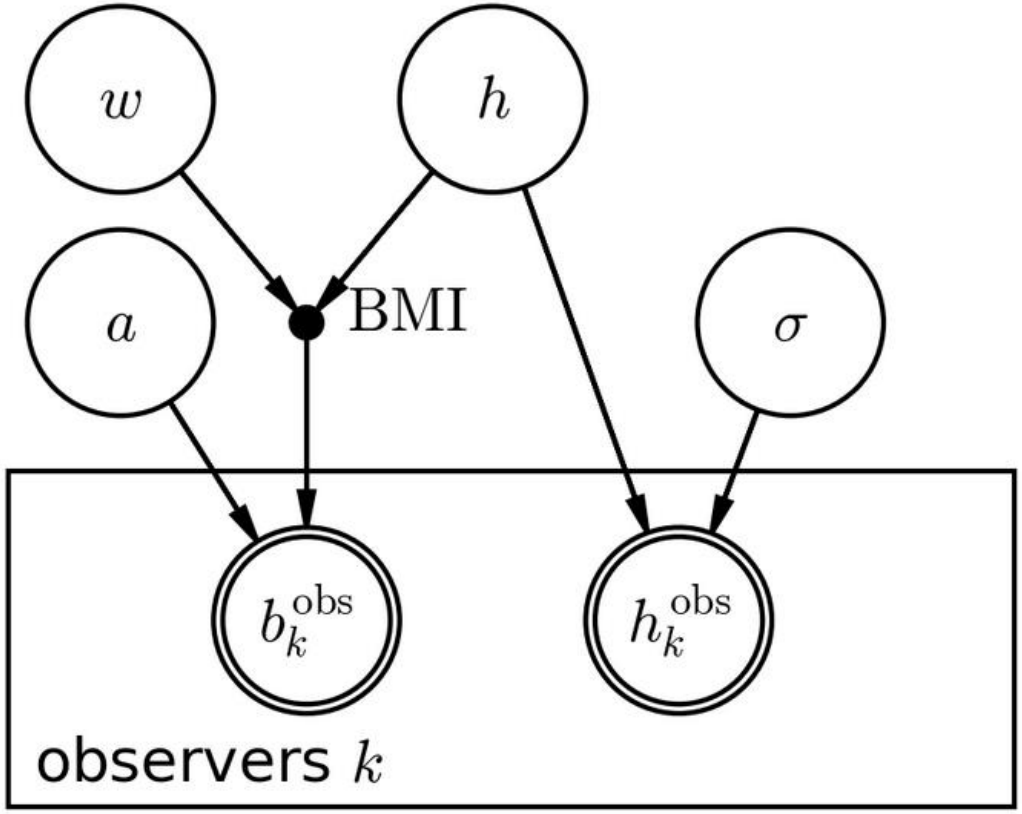
- $BMI = \frac{weight}{height^2}$ (in SI units)
- Relation between obesity and BMI

$\Pr(b|\text{BMI}) = U[\alpha, \beta]$, where:

α	β	b
12.1	18.5	"Underweight"
18.5	24.9	"Normal"
24.9	31.4	"Overweight"

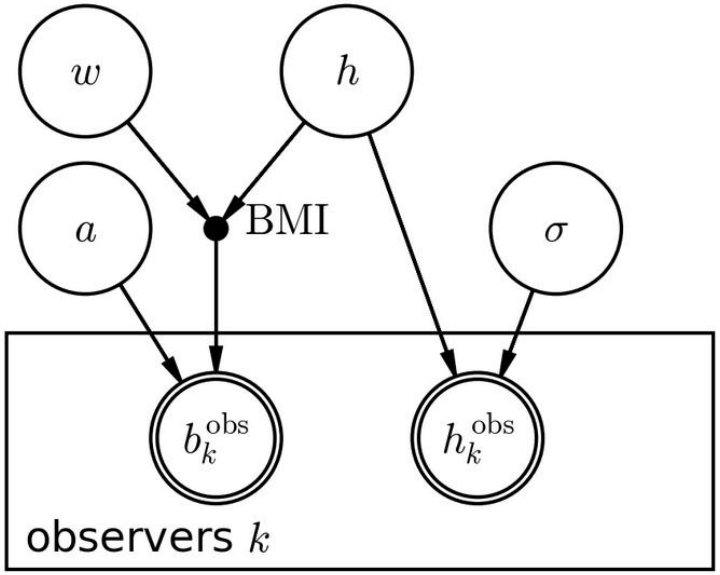
PGM of the first method

- w: Phil's weight
- h: Phil's height
- b_k: observed obesity
- h_k: observed height
- σ: error of height estimate
- a: accuracy of determining the correct obesity



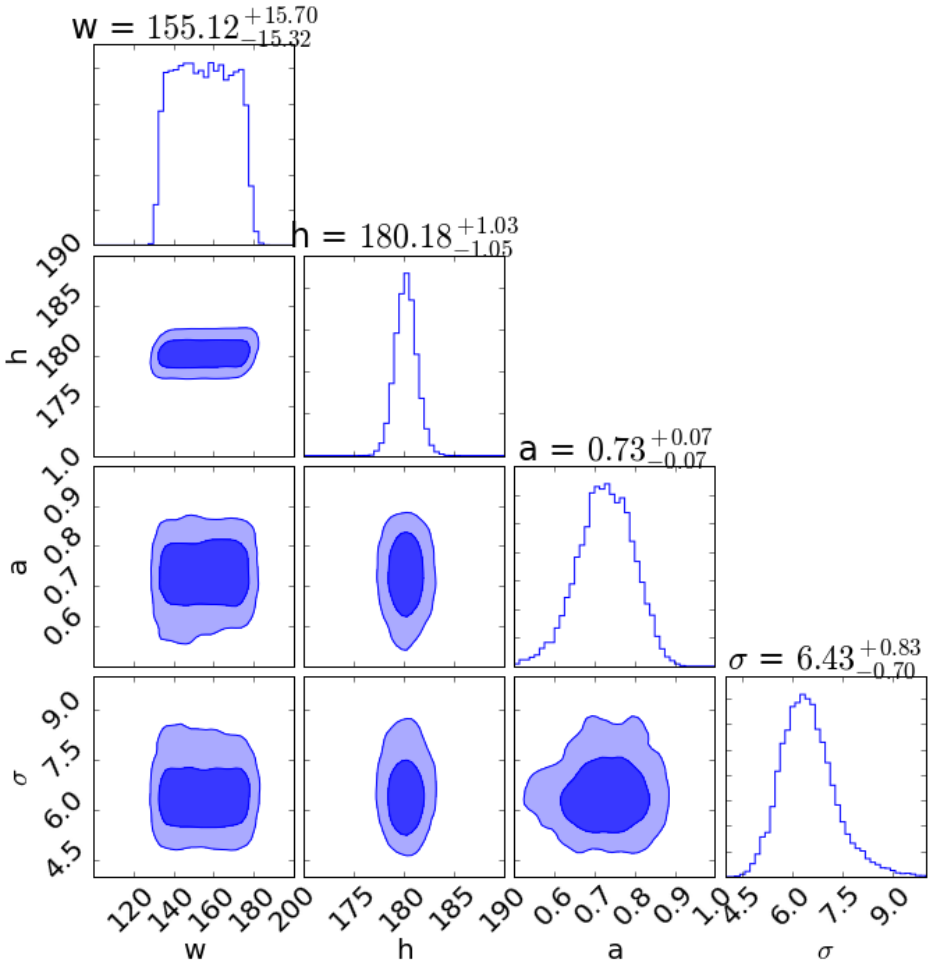
Priors and Likelihoods

- $P(h_k|h,\sigma) = Normal(h_k|h,\sigma)$
- $P(b_k|BMI) = \begin{cases} a \text{ if BMI in the correct range} \\ \frac{1-a}{2} \text{ if not} \end{cases}$
- a has flat prior in [0.5,1]
- w, h, σ also have flat priors



Simple MCMC result

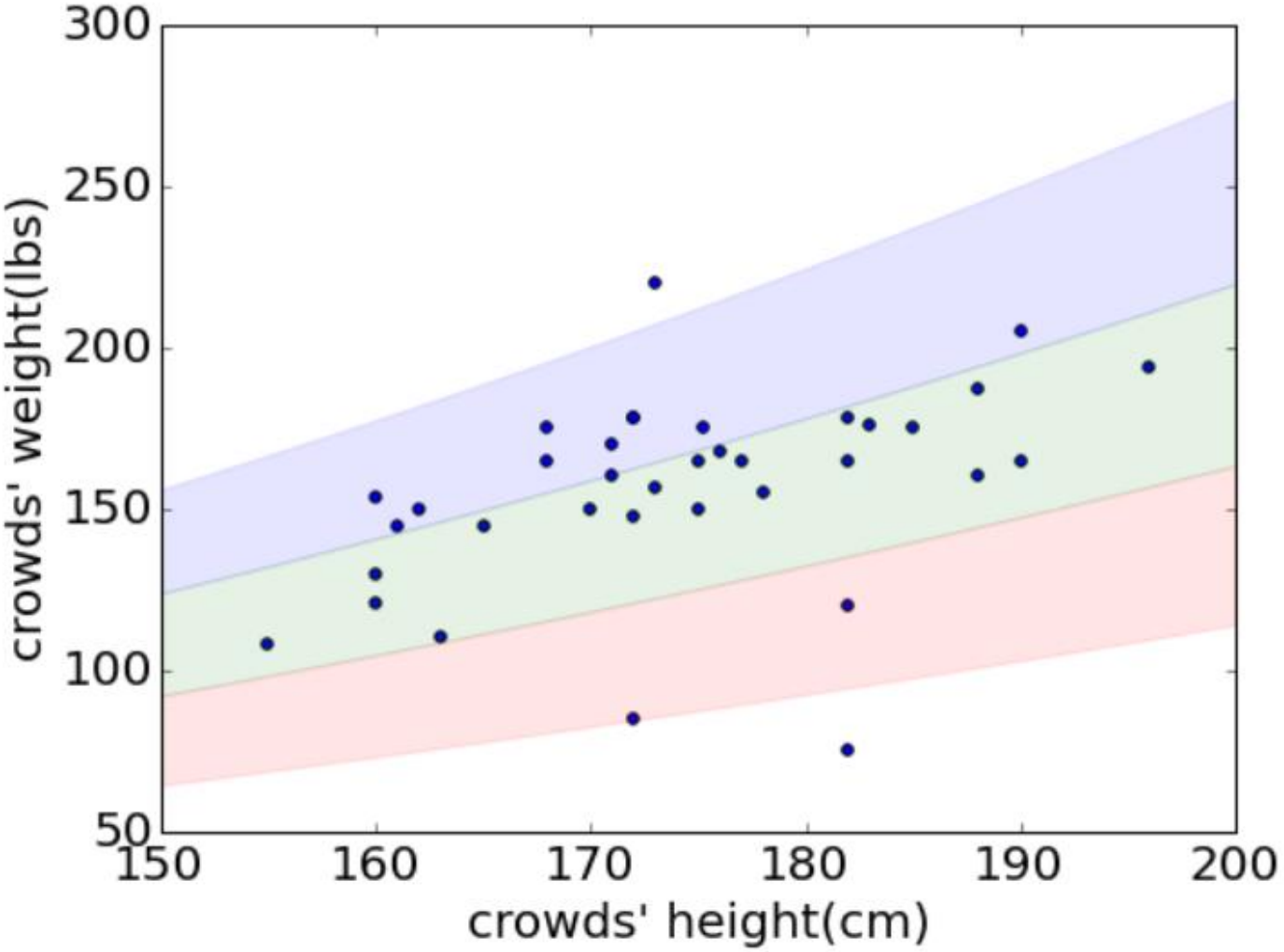
- Phil's weight = 155 ± 15 (lbs)



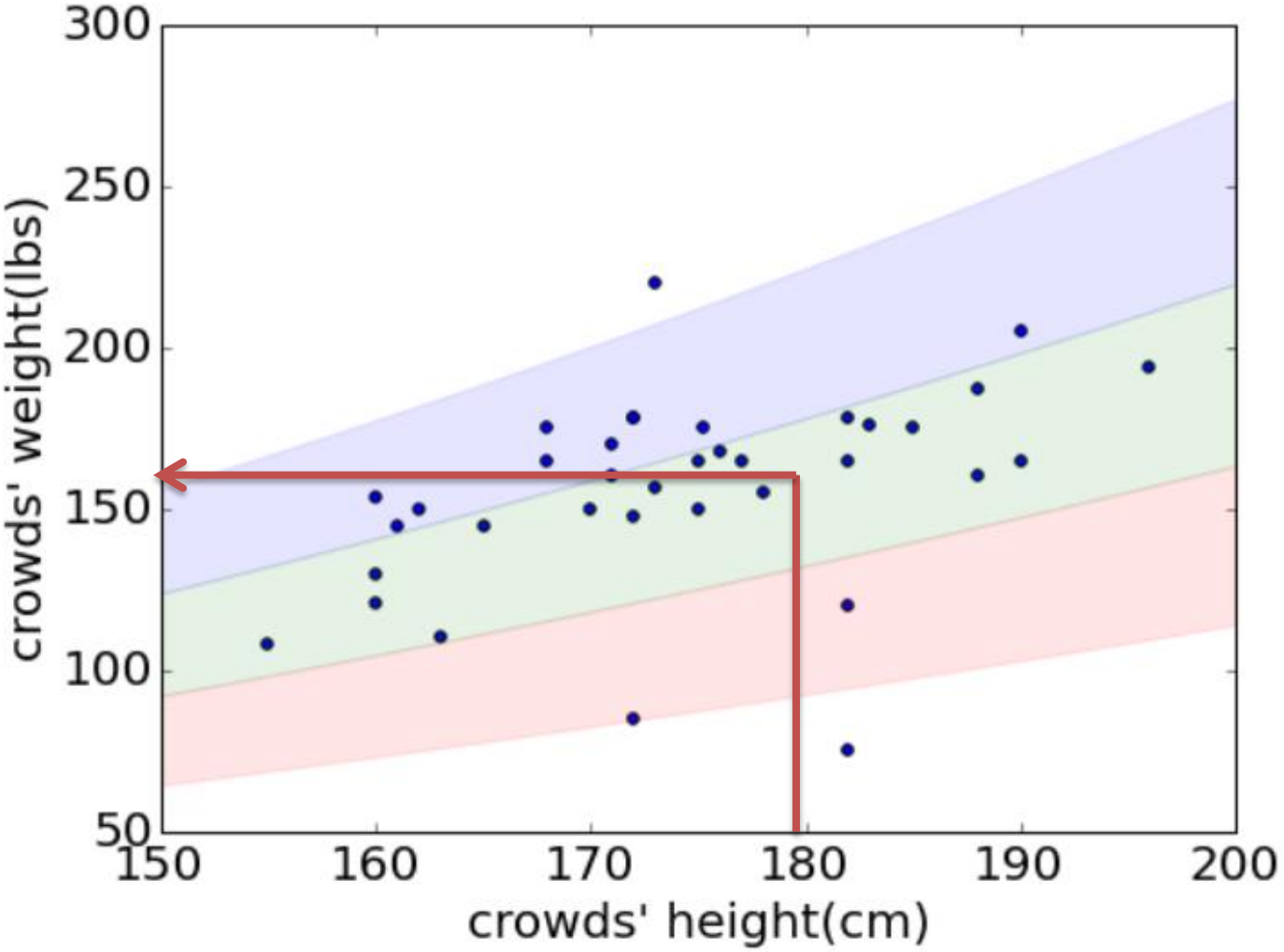
Second Method –
Assume Phil is drawn from the same distribution of the crowd

- pa: crowds' own age
- ph: crowds' own height
- pw: crowds' own weight
- a: crowds' estimates of Phil's age
- h: crowds' estimates of Phil's h
- **Assume Phil's age, height, weight has the same distribution as the crowds'**

Distribution of crowds' height and weight

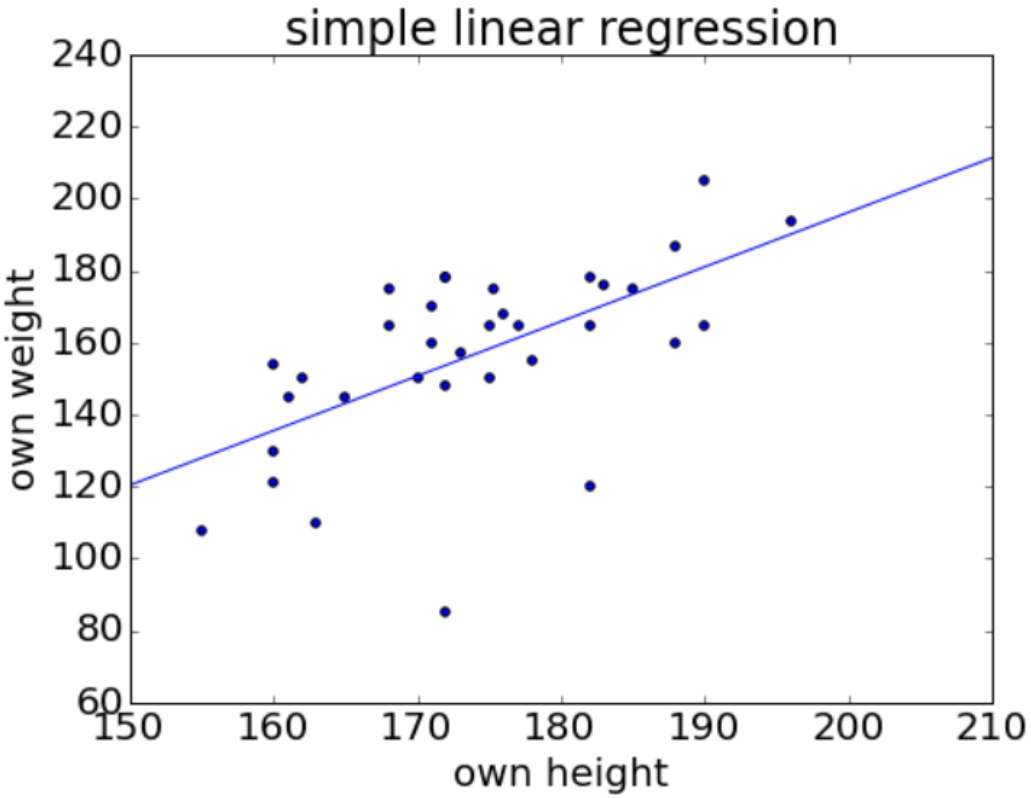


Distribution of crowds' height and weight



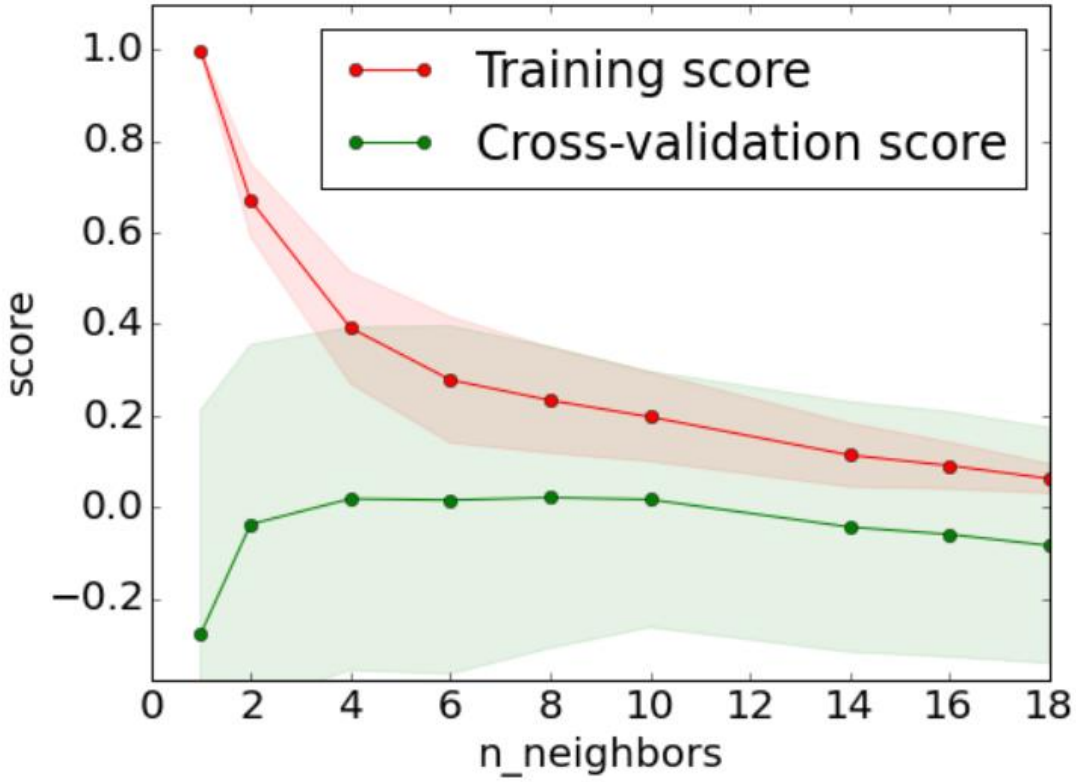
Result of linear regression

  - Phil's height = 180 ± 1 (cm)
- Phil's weight = 166 ± 20 (lbs)



Machine learning
(age, height) ⇒ weight

- Use linear model and KNN model
  - R^2 score of linear model: 0.15 ± 0.33



Machine learning results

Phil's age = 41 ± 1 (years)

Phil's height = 180 ± 1 (cm)

- Linear model:
  - Phil's weight = 170 (lbs)
- KNN (with k = 8):
  - Phil's weight = 158 (lbs)
- Possible ways to get uncertainty:
  - Use results from different training sets

Comparison of results

- Bayesian: 155 ± 15 (lbs)
- Linear regression: 166 ± 20 (lbs)
- Machine learning:
  - Linear model: 170 (lbs)
  - KNN: 158 (lbs)