

The confection



m&m's
(185g)



Jelly Belly
(100g)



Chocolate Raisins
(200g)

The importance of guessing

<http://StreetFightingMath.com/>

Stuff Inf2b students wrote

Number M&Ms: ~~185~~ 204
 Number Jelly Belly: ~~146~~ 146
 Num. choc-raisin blobs: ~~87~~ 87

Number M&Ms: ~~185~~ 185
 Number Jelly Belly: ~~8~~ 180
 Num. choc-raisin blobs: ~~190~~ 190

Number M&Ms: ~~240~~ 240
 Number Jelly Belly: ~~150~~ 150
 Num. choc-raisin blobs: ~~130~~ 130

Number M&Ms: ~~247~~ 247
 Number Jelly Belly: ~~75~~ 75
 Num. choc-raisin blobs: ~~89~~ 89

Number M&Ms: ~~70~~ 70
 Number Jelly Belly: ~~83~~ 83
 Num. choc-raisin blobs: ~~100~~ 100

Number M&Ms: ~~150~~ 150 152 202 82
 Number Jelly Belly: ~~70~~ 72
 Num. choc-raisin blobs: ~~150~~ 132 102

Number M&Ms: ~~168~~ 168
 Number Jelly Belly: ~~98~~ 98
 Num. choc-raisin blobs: ~~139~~ 139

Number M&Ms: ~~84~~ 84
 Number Jelly Belly: ~~52~~ 52
 Num. choc-raisin blobs: ~~133~~ 133

F83D M3

Number M&Ms: 90
 Number Jelly Belly: 40
 Num. choc-raisin blobs: ~~40~~
 or pose likely the average of all other guesses...
 Full name:
 (to award prize only)

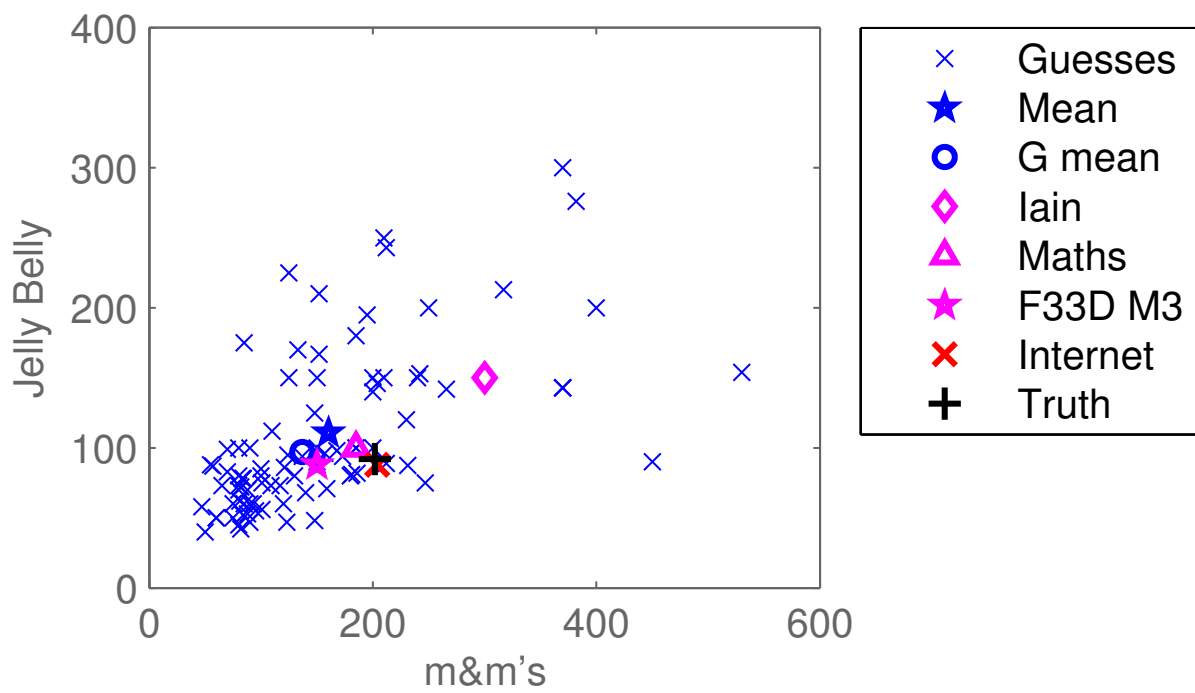
Number M&Ms: 231.25
 Number Jelly Belly: 87.5
 Num. choc-raisin blobs: 133.34
 Full name: ANON
 (to award prize only)

$$\rho = 1 \frac{g}{cm^3}, \quad \rho = 1.7 \frac{g}{cm^3}$$

$$.5 \text{ cm}^3 \text{ each}$$

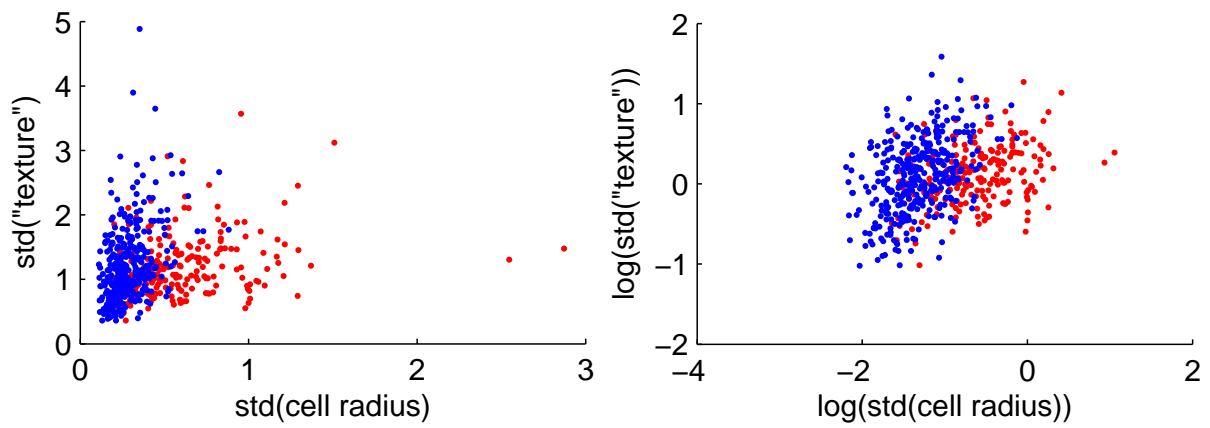
$$\rho = \frac{m}{V} \Rightarrow m = \rho V = 1.5 \cdot .5 = .75$$

A 2D space



For 3D and more, check out the code on the website.

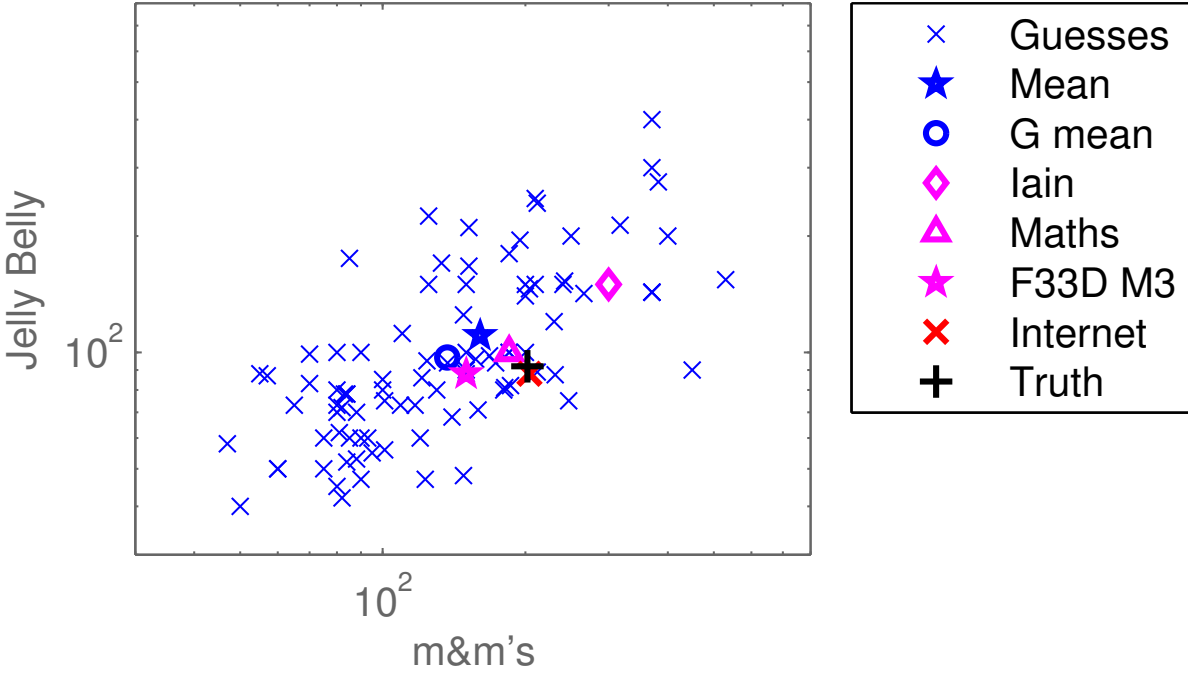
Often log-transform +ve data



Wisconsin breast cancer data

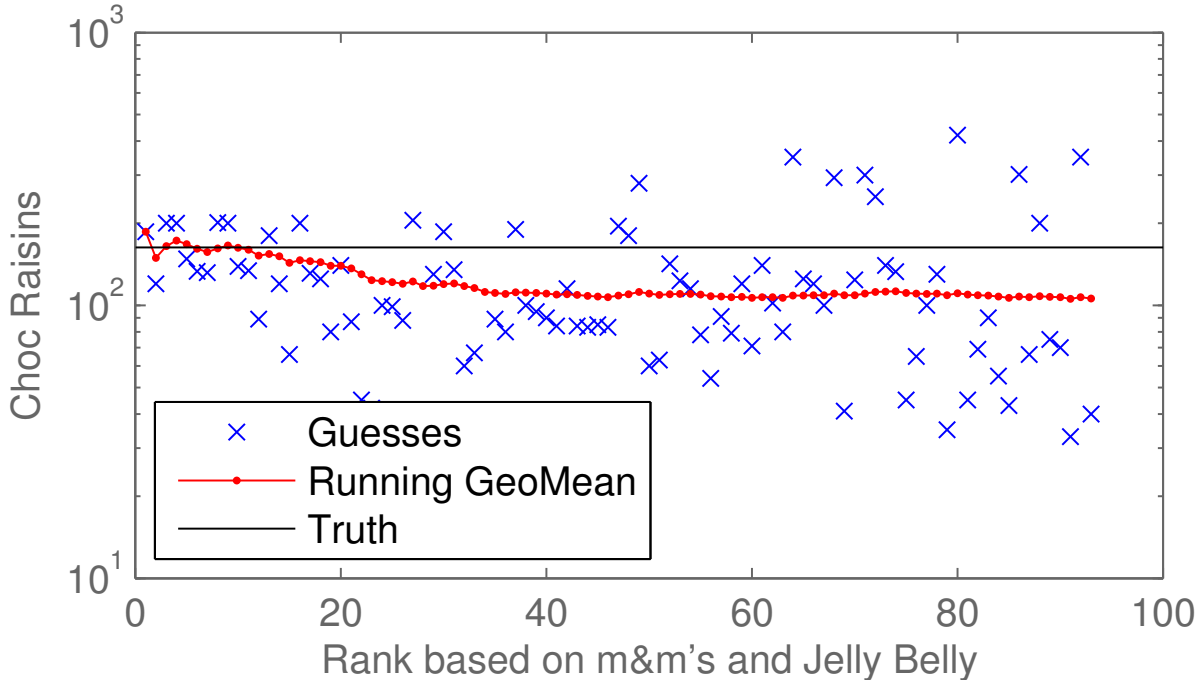
UCI ML repository

Count guesses on log-scale



Were some people just lucky?

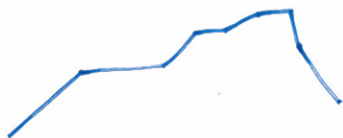
Ranking by past performance



Ensemble of Models

Ensembles can ^① reduce overfitting
and. can ^② reduce underfitting

Build complicated f^N



from
simple
pieces

Eg of ①

Bayesian model averaging:

$$p(y | \underline{x}, D) = \int p(y | \underline{x}, \underline{w}) p(\underline{w} | D) d\underline{w}$$

$$\approx \frac{1}{S} \sum_{s=1}^S p(y | \underline{x}, \underline{w}^{(s)}) \quad \underline{w}^{(s)} \sim p(\underline{w} | D)$$

↑
Ensemble of S predictors

What about making point prediction or guesses?

Kaggle give you a loss fn

$$L(\hat{y}; y) \stackrel{\text{eg}}{=} (\hat{y} - y)^2 \quad \text{or} \quad |\hat{y} - y| \\ \uparrow \quad \quad \uparrow \\ \text{guess} \quad \text{answer} \quad \quad \text{or} \dots$$

Then minimize expected loss:

$$\begin{aligned} \text{Minimize } \mathbb{E}_{p(y|\underline{x}, D)} [L(\hat{y}; y)] \\ = \int p(y|\underline{x}, D) L(\hat{y}; y) dy \end{aligned}$$

For square error:

$$\Rightarrow \hat{y} = \text{mean of } p(y|\underline{x}, D) \\ \text{or } \mathbb{E}_{p(y|\underline{x}, D)} [y]$$

For absolute error:

$$\Rightarrow \hat{y} = \text{median of } p(y|\underline{x}, D)$$

^{was}
Another Lot of averaging predictions: Bagging

Bootstrap aggregation

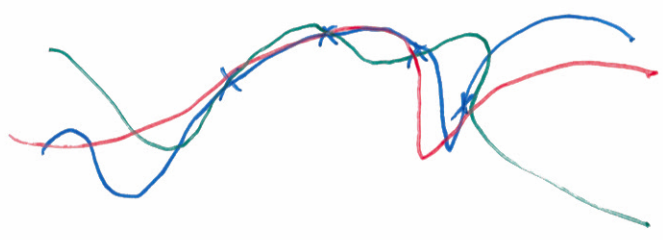
You have a dataset of N examples
Train time

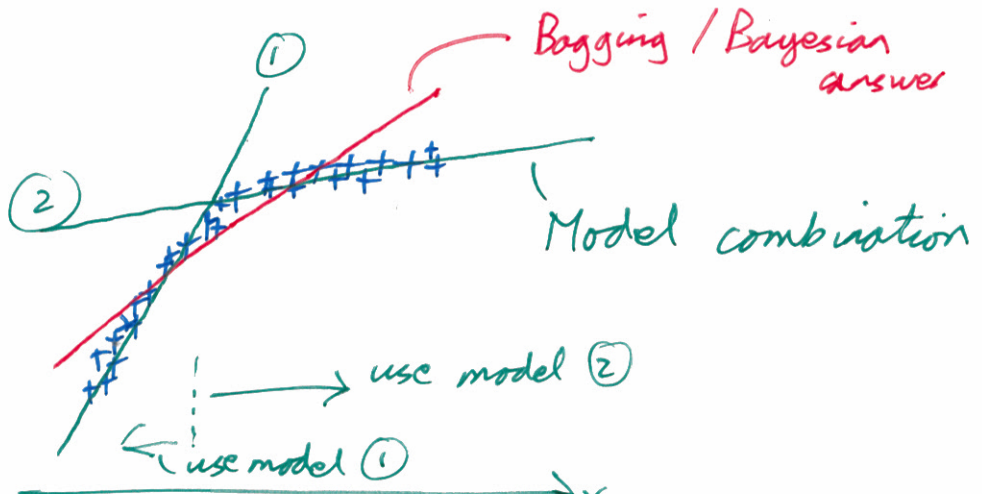
for $s = 1 \dots S$

Create a new dataset of N examples
by sampling with replacement from
training set. (Bootstrap)

Test time Fit your model \rightarrow predictor s

Average predictions of S models





Model Combination

$$p(y|x, \theta) = \sum_k p(y|x, z=k, \theta) p(z=k|x, \theta)$$

\uparrow params \uparrow $k \in \{1, 2, \dots\}$

predictor/experts classifier

"Mixture of experts"

Fit θ , Max. Likelihood, + regularization.

Bayesian \rightarrow Laplace / Var approx of $p(\theta|D)$
 \rightarrow Or Bagging.

I showed you Bucilă et al. and Caruana et al. papers.
 See links in notes.