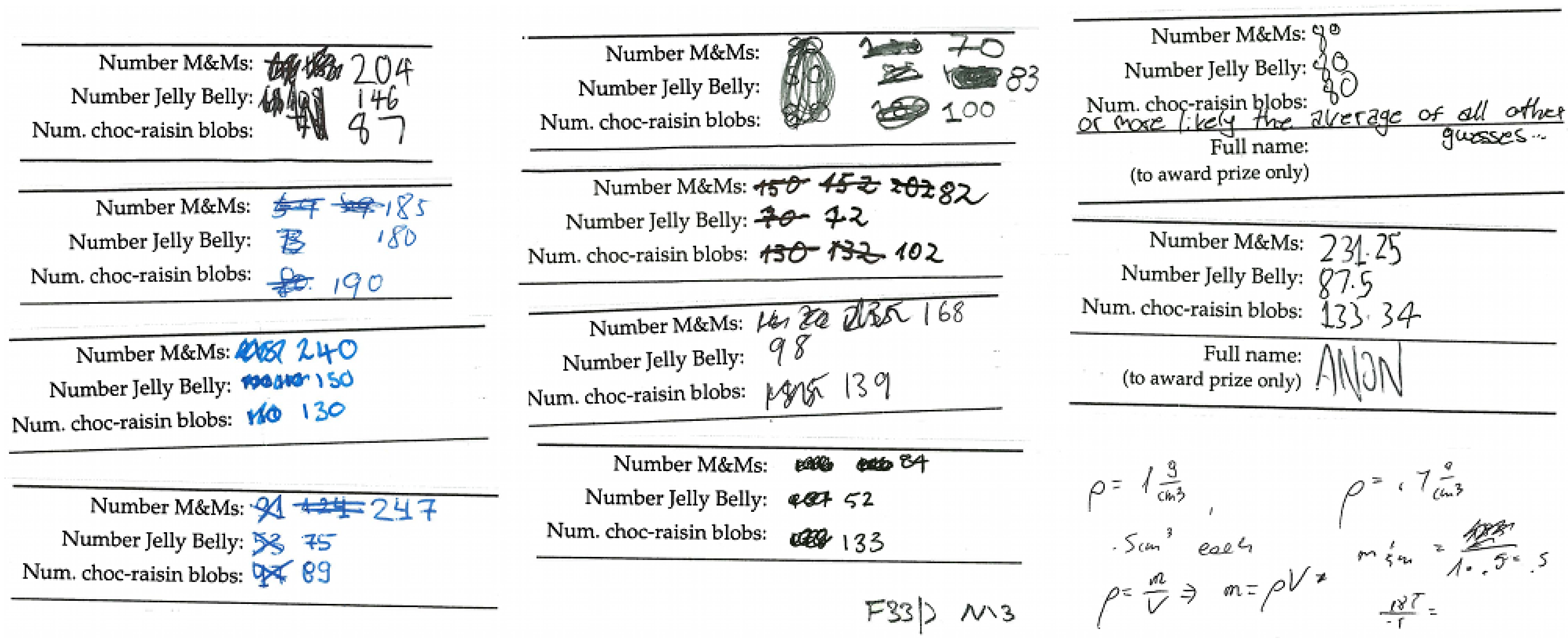
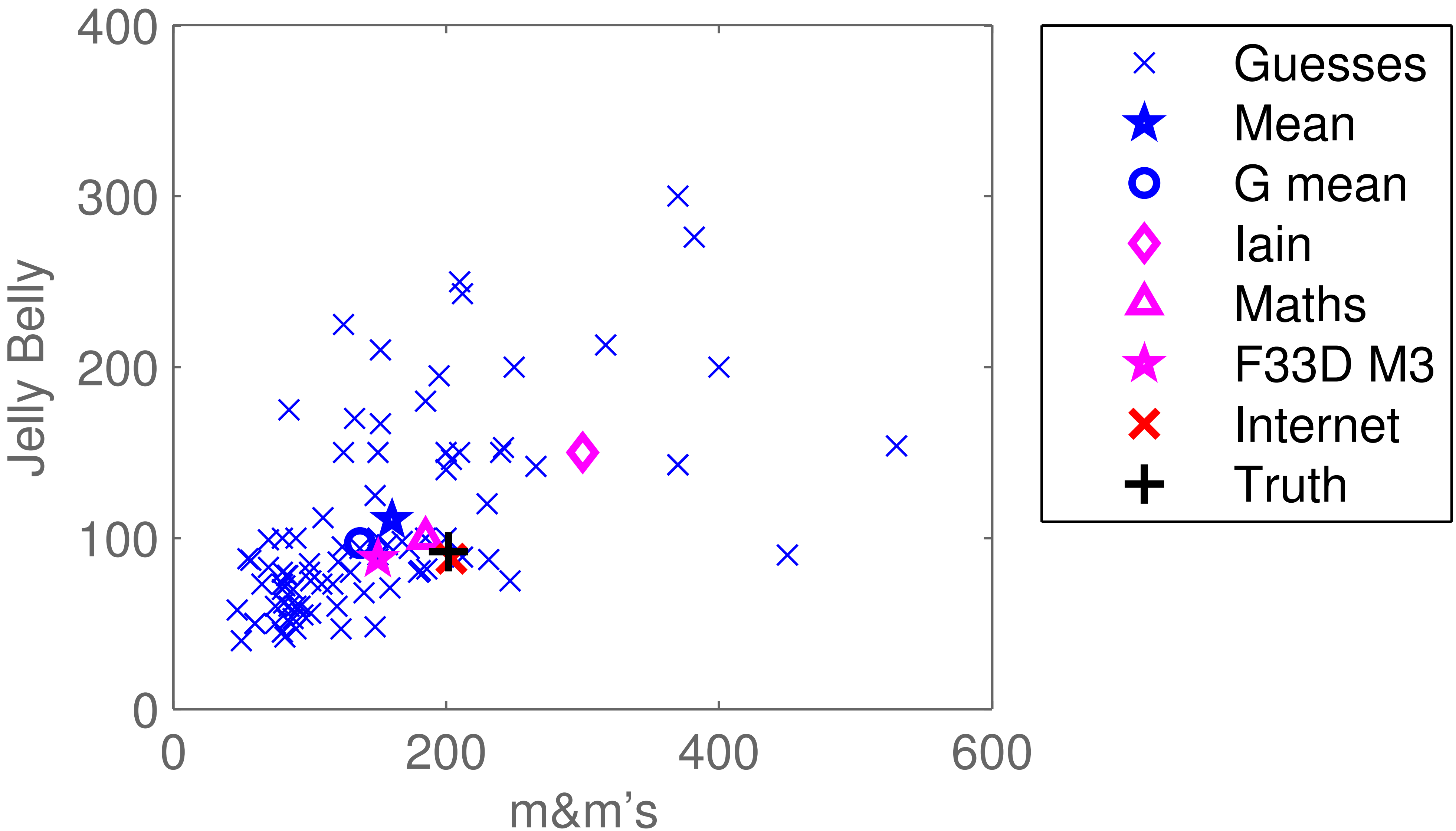
The confection

m&m's	Jelly Belly	Chocolate Raisins
(185g)	(100g)	(200g)

Stuff Inf2b students wrote



A 2D space



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

The importance of guessing

`http://StreetFightingMath.com/`

Dimensional Analysis

$$C = A\,B$$
$N \times D \quad N \times K \quad K \times D$

$\rightarrow A \searrow$
$\qquad\qquad C \rightarrow \cdots \rightarrow z$ (cost)
$\rightarrow B \nearrow$

$$\bar{C}_{ij} = \frac{\partial z}{\partial C_{ij}}$$

$$\bar{A}_{ij} = \frac{\partial z}{\partial A_{ij}}$$

(1) Physical units

$$\underset{m\cdot kg}{c} = \underset{m}{a}\,\underset{kg}{b}$$

Units of $\frac{\partial z}{\partial a}$: $\frac{\text{cost}}{m}$

$\bar{c} = \frac{\partial z}{\partial c}$: $\frac{\text{cost}}{m\cdot kg}$

$$\bar{a} = \frac{\partial z}{\partial a} = \frac{\partial z}{\partial c}\frac{\partial c}{\partial a}$$

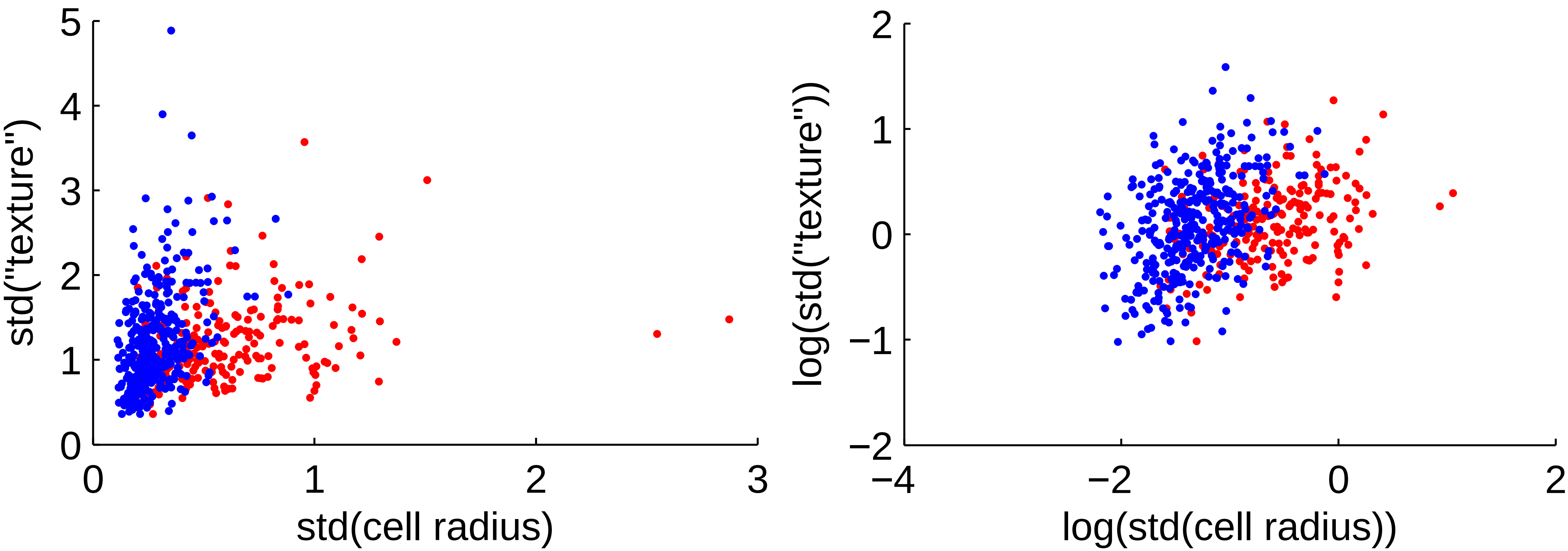
$$= \bar{c}\,b$$

(2) Match sizes of arrays

$$\bar{A} = \bar{C}\,B^T \;?$$
$N \times K \quad N \times D \quad D \times K$

$\rightarrow$ write test case.

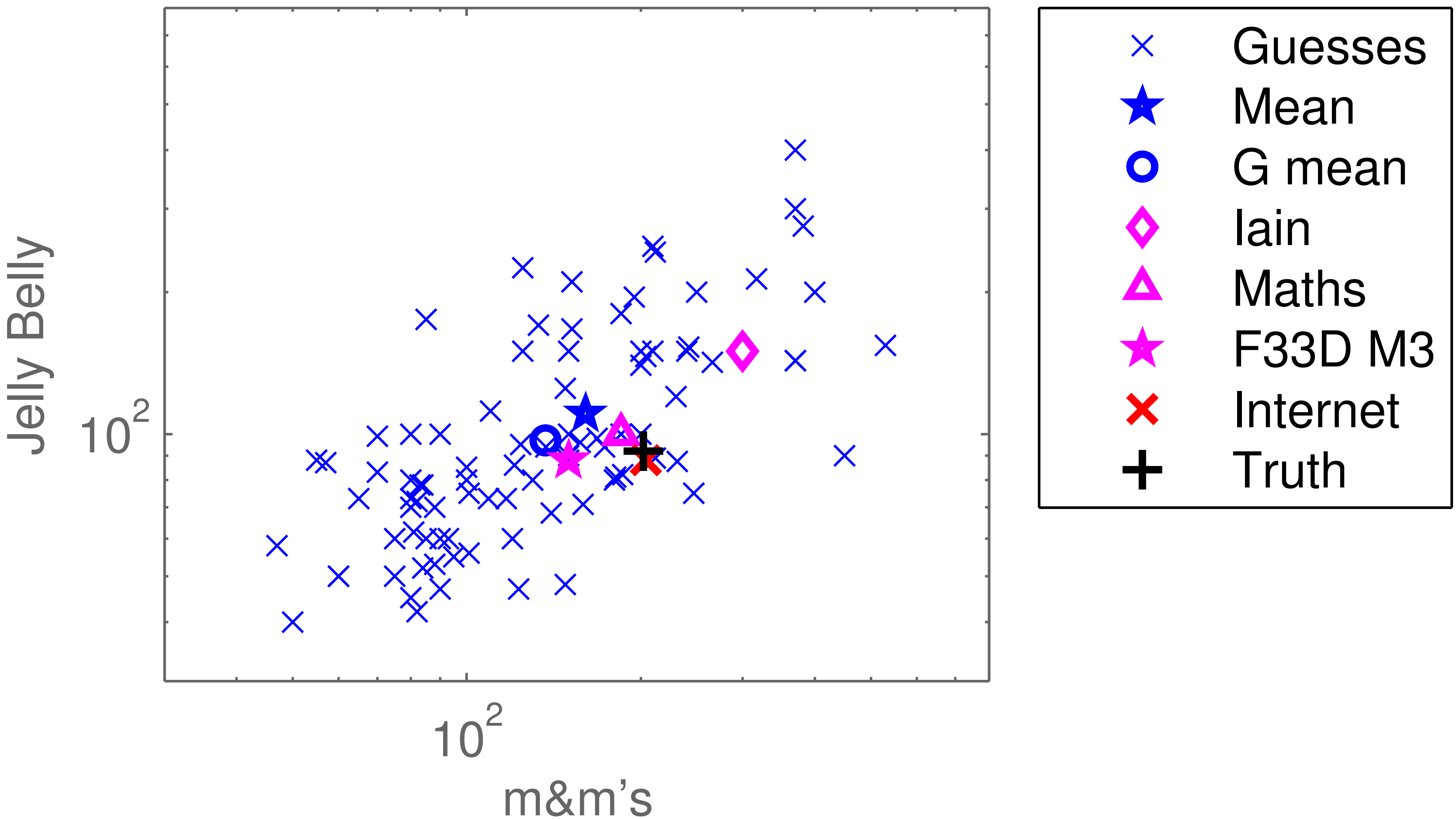
Often log-transform +ve data



Wisconsin breast cancer data

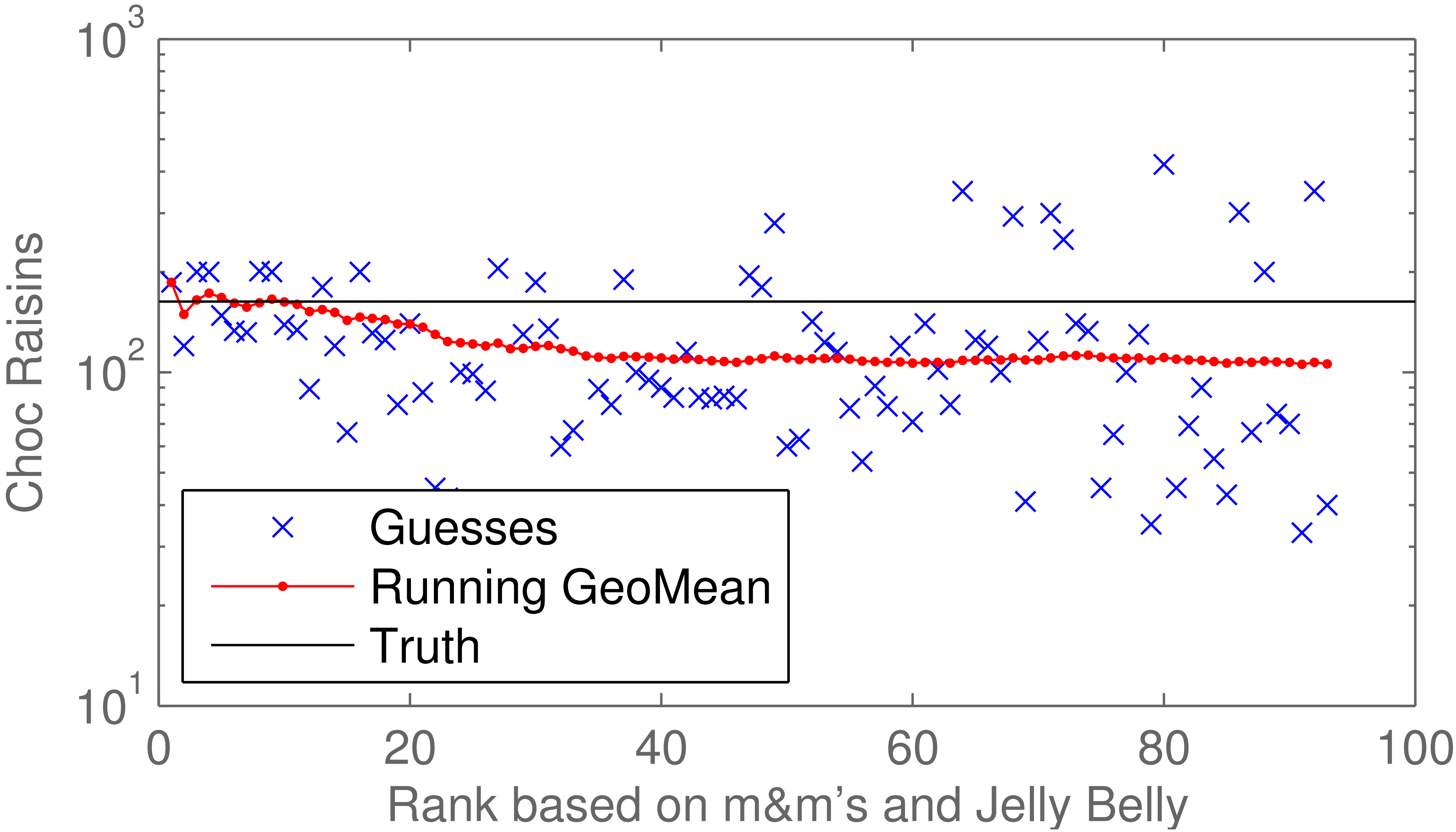
UCI ML repository

Count guesses on log-scale



Were some people just lucky?

Ranking by past performance



Ensemble of Models

Two motivations:

1. Reduce over-fitting
2. Reduce under-fitting

Example 1

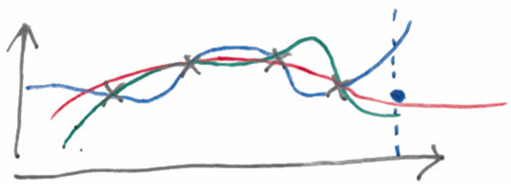
Bayesian model averaging:

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

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

$$\underline{w}^{(s)} \sim p(\underline{w} \mid D)$$

↑ Ensemble of S predictors.



Another way of ensembling: Bagging

Bootstrap aggregation

You have N training examples

Training time:

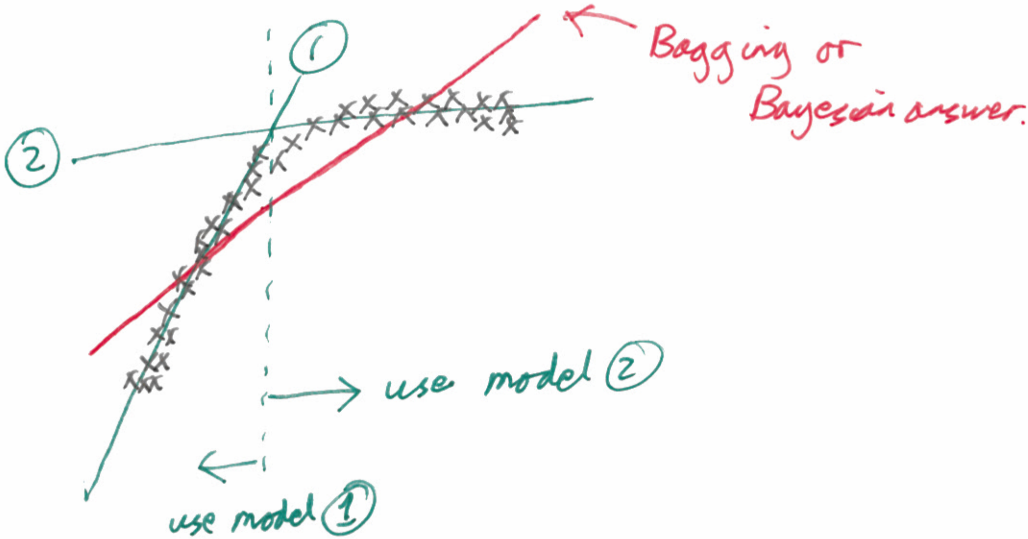
for $s = 1 \dots S$:

Bootstrap: create a new dataset with N datapoints sampled with replacement from training data.

Fit your model → predictor s

Test time:

Average predictions.



2 Model Combination

$s \in \{1, 2 \ldots S\}$

$$p(y \mid \underline{x}, \theta) = \sum_{s=1}^{S} \underbrace{p(y \mid \underline{x}, z=s, \theta)}_{\text{Any regression model}} \underbrace{p(z=s \mid \underline{x}, \theta)}_{\text{classifier}}$$

"Mixture of experts"

→ Fit θ, regularize

Bayesian, Bagging,

For the papers I mentioned, see the typeset notes.