

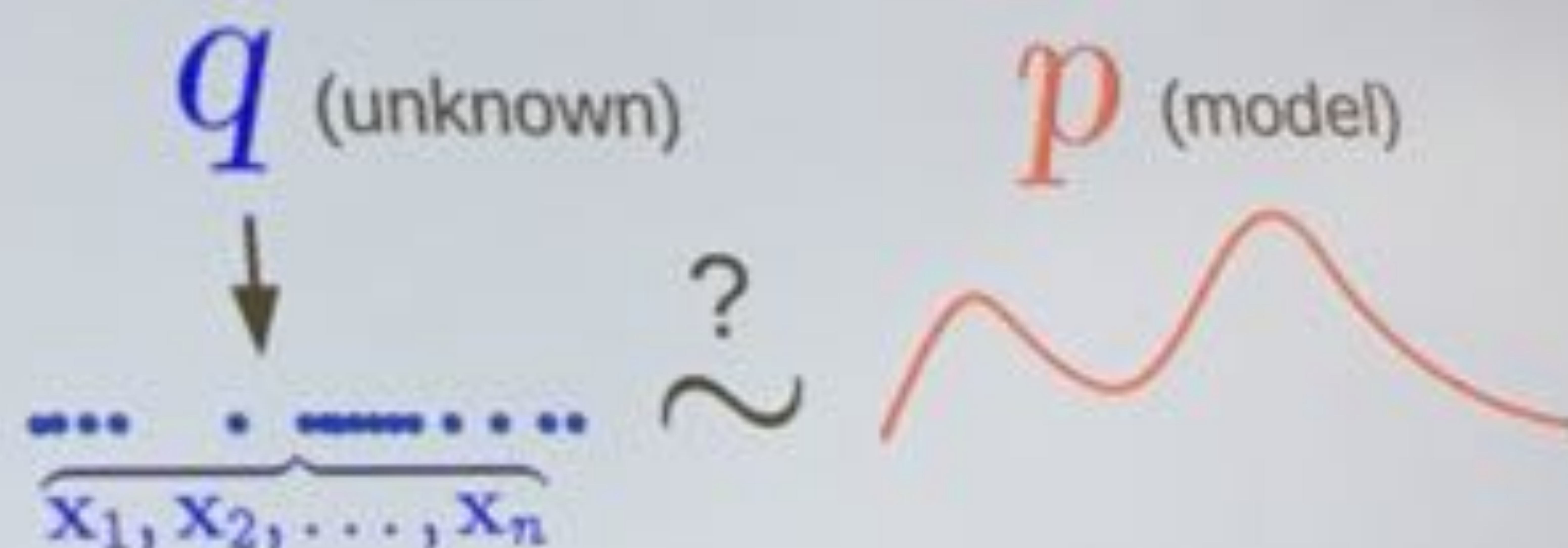
Model Criticism



Goals:

- 1 Test if a (complicated) **model** fits the **data**.
- 2 If it does not, show **a location** where it fails.

Problem Setting: Goodness-of-Fit Test



Test goal: Are **data** from the **model** p ?

- 1 **Nonparametric.**
- 2 **Linear-time.** Runtime is $\mathcal{O}(n)$. Fast.
- 3 **Interpretable.** Model criticism by finding \star .

Model Criticism by Maximum Mean Discrepancy [Gretton et al., 2012]

- Find a location \mathbf{v} at which q and p differ most [Jitkrittum et al., 2016].

$$\text{witness}(\mathbf{v}) = \mathbb{E}_{\mathbf{x} \sim q} [k_{\mathbf{v}}(\mathbf{x})] - \mathbb{E}_{\mathbf{y} \sim p} [k_{\mathbf{v}}(\mathbf{y})]$$

Model Criticism by Maximum Mean Discrepancy [Gretton et al., 2012]

- Find a location \mathbf{v} at which q and p differ most [Jitkrittum et al., 2016].

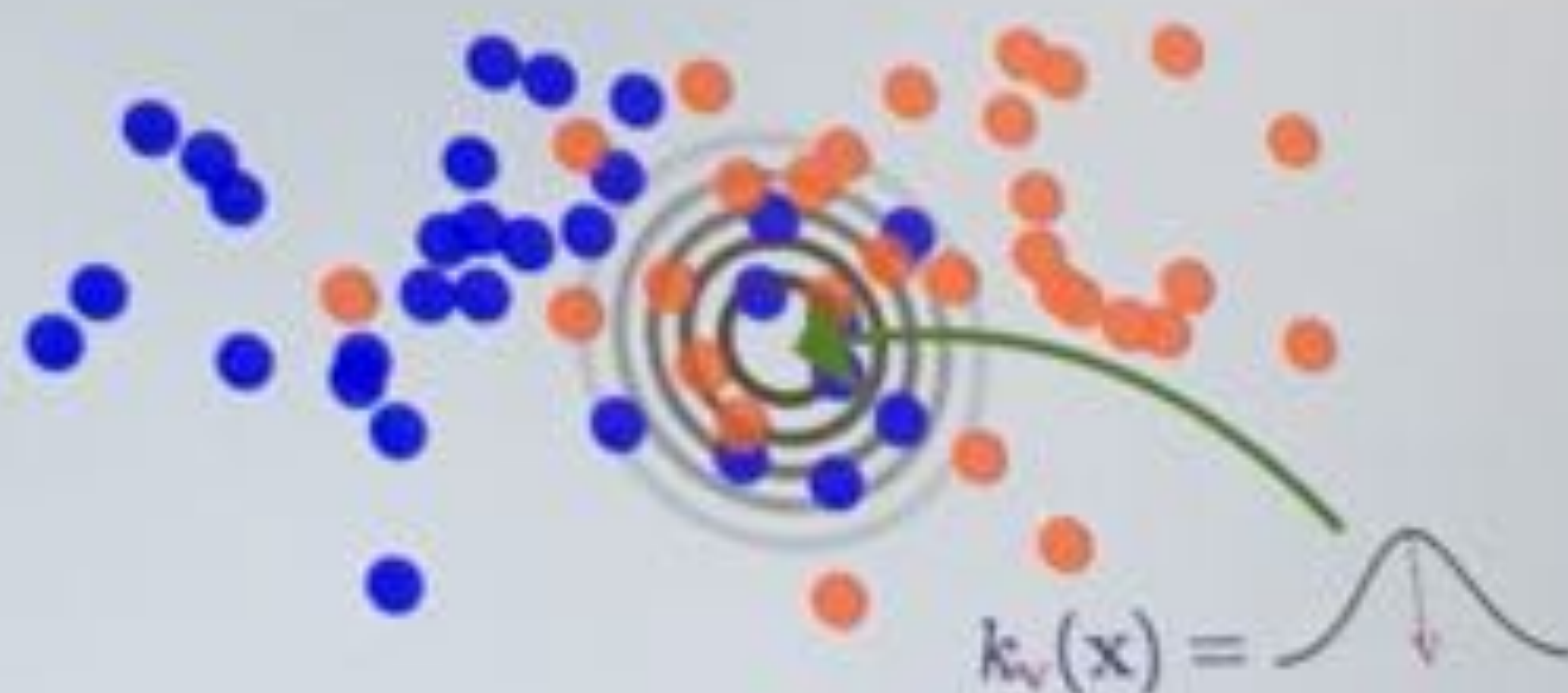
$$\text{witness}(\mathbf{v}) = \mathbb{E}_{\mathbf{x} \sim q} \left[\text{curve}_q(\mathbf{v}) \right] - \mathbb{E}_{\mathbf{y} \sim p} \left[\text{curve}_p(\mathbf{v}) \right]$$


$$\text{score}(\mathbf{v}) = \frac{|\text{witness}(\mathbf{v})|}{\text{standard deviation}(\mathbf{v})}$$

Model Criticism by Maximum Mean Discrepancy [Gretton et al., 2012]

- Find a location \mathbf{v} at which q and p differ most [Jitkrittum et al., 2016].

score: 0.008



$$\text{witness}(\mathbf{v}) = \mathbb{E}_{\mathbf{x} \sim q}[\text{Gaussian curve}] - \mathbb{E}_{\mathbf{y} \sim p}[\text{Gaussian curve}]$$

$$\text{score}(\mathbf{v}) = \frac{|\text{witness}(\mathbf{v})|}{\text{standard deviation}(\mathbf{v})}$$

Model Criticism by Maximum Mean Discrepancy [Gretton et al., 2012]

- Find a location \mathbf{v} at which q and p differ most [Jitkrittum et al., 2016].

score: 25



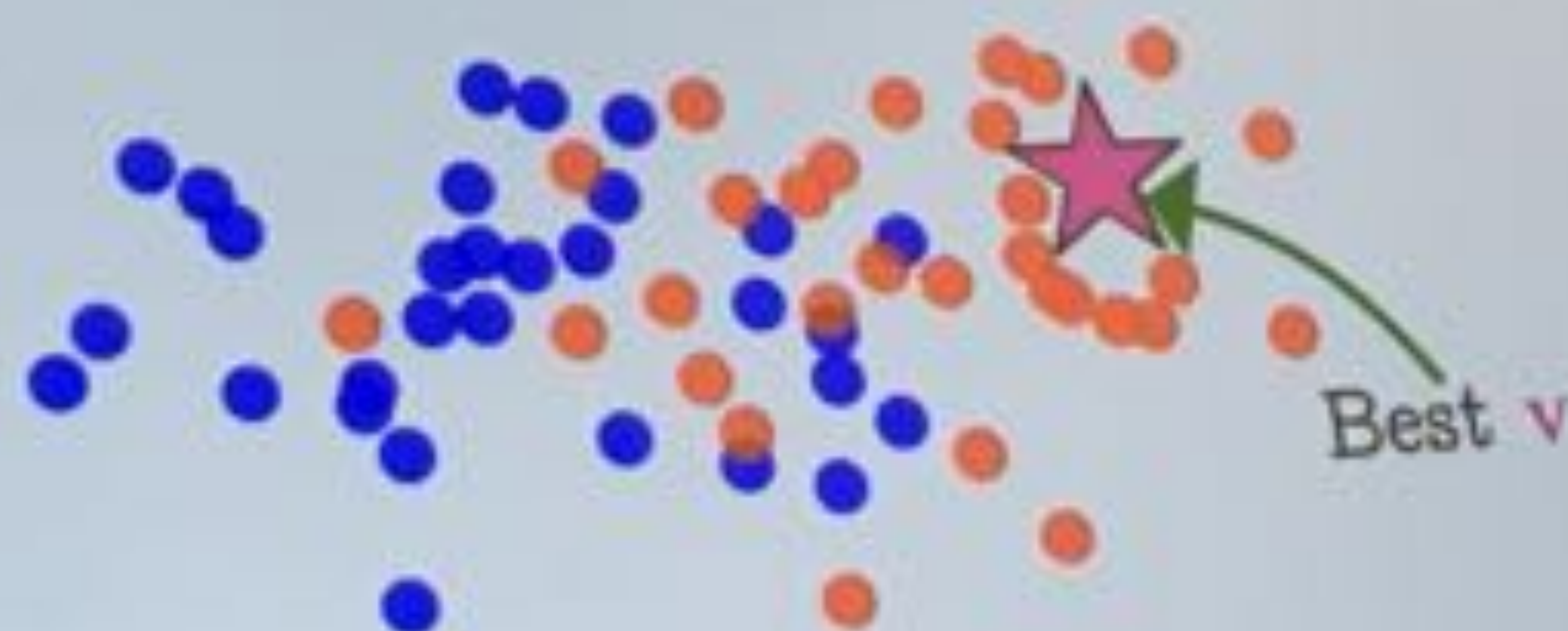
$$\text{witness}(\mathbf{v}) = \mathbb{E}_{\mathbf{x} \sim q} \left[\text{kernel}(\mathbf{x}, \mathbf{v}) \right] - \mathbb{E}_{\mathbf{y} \sim p} \left[\text{kernel}(\mathbf{y}, \mathbf{v}) \right]$$

$$\text{score}(\mathbf{v}) = \frac{|\text{witness}(\mathbf{v})|}{\text{standard deviation}(\mathbf{v})}$$

Model Criticism by Maximum Mean Discrepancy [Gretton et al., 2012]

- Find a location \mathbf{v} at which q and p differ most [Jitkritum et al., 2016].

score: 25



$$\text{witness}(\mathbf{v}) = \mathbb{E}_{\mathbf{x} \sim q} \left[\text{kernel}(\mathbf{x}, \mathbf{v}) \right] - \mathbb{E}_{\mathbf{y} \sim p} \left[\text{kernel}(\mathbf{y}, \mathbf{v}) \right]$$
$$\text{score}(\mathbf{v}) = \frac{|\text{witness}(\mathbf{v})|}{\text{standard deviation}(\mathbf{v})}$$

No sample from p .
Difficult to generate.

Model Criticism by Maximum Mean Discrepancy [Gretton et al., 2012]

- Find a location \mathbf{v} at which q and p differ most [Jitkrittum et al., 2016].

score: 25



$$\text{witness}(\mathbf{v}) = \mathbb{E}_{\mathbf{x} \sim q} \left[\text{bell curve} \right] - \mathbb{E}_{\mathbf{y} \sim p} \left[\text{bell curve} \right]$$



The diagram shows two bell curves, one for q and one for p . A vertical dashed line from the peak of the q curve is labeled \mathbf{v} . Another vertical dashed line from the peak of the p curve is also labeled \mathbf{v} . A red arrow points from the text "No sample from p . Difficult to generate." to the \mathbf{v} label on the p curve.

$$\text{score}(\mathbf{v}) = \frac{|\text{witness}(\mathbf{v})|}{\text{standard deviation}(\mathbf{v})}$$

No sample from p .
Difficult to generate.

The Stein Witness Function [Liu et al., 2016, Chwialkowski et al., 2016]

Problem: No sample from p . Cannot estimate $\mathbb{E}_{y \sim p}[k_v(y)]$.

$$(\text{Stein}) \text{ witness}(v) = \mathbb{E}_{x \sim q}[T_p \text{ ] - \mathbb{E}_{y \sim p}[T_p \text{ ]$$

The Stein Witness Function [Liu et al., 2016, Chwialkowski et al., 2016]

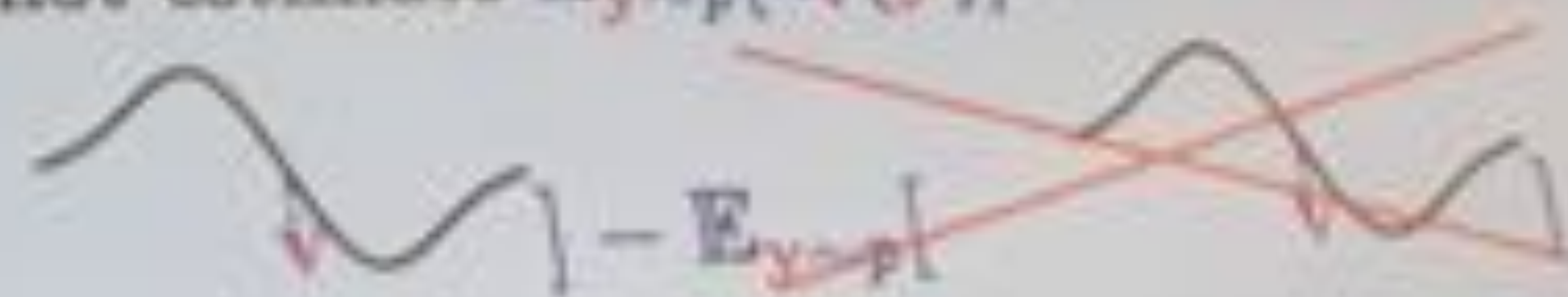
Problem: No sample from p . Cannot estimate $E_{y \sim p}[k_v(y)]$.

$$(\text{Stein}) \text{ witness}(v) = E_{x \sim q}[\text{witness}(v, x)] - E_{y \sim p}[\text{witness}(v, y)]$$

The Stein Witness Function [Liu et al., 2016, Chwialkowski et al., 2016]

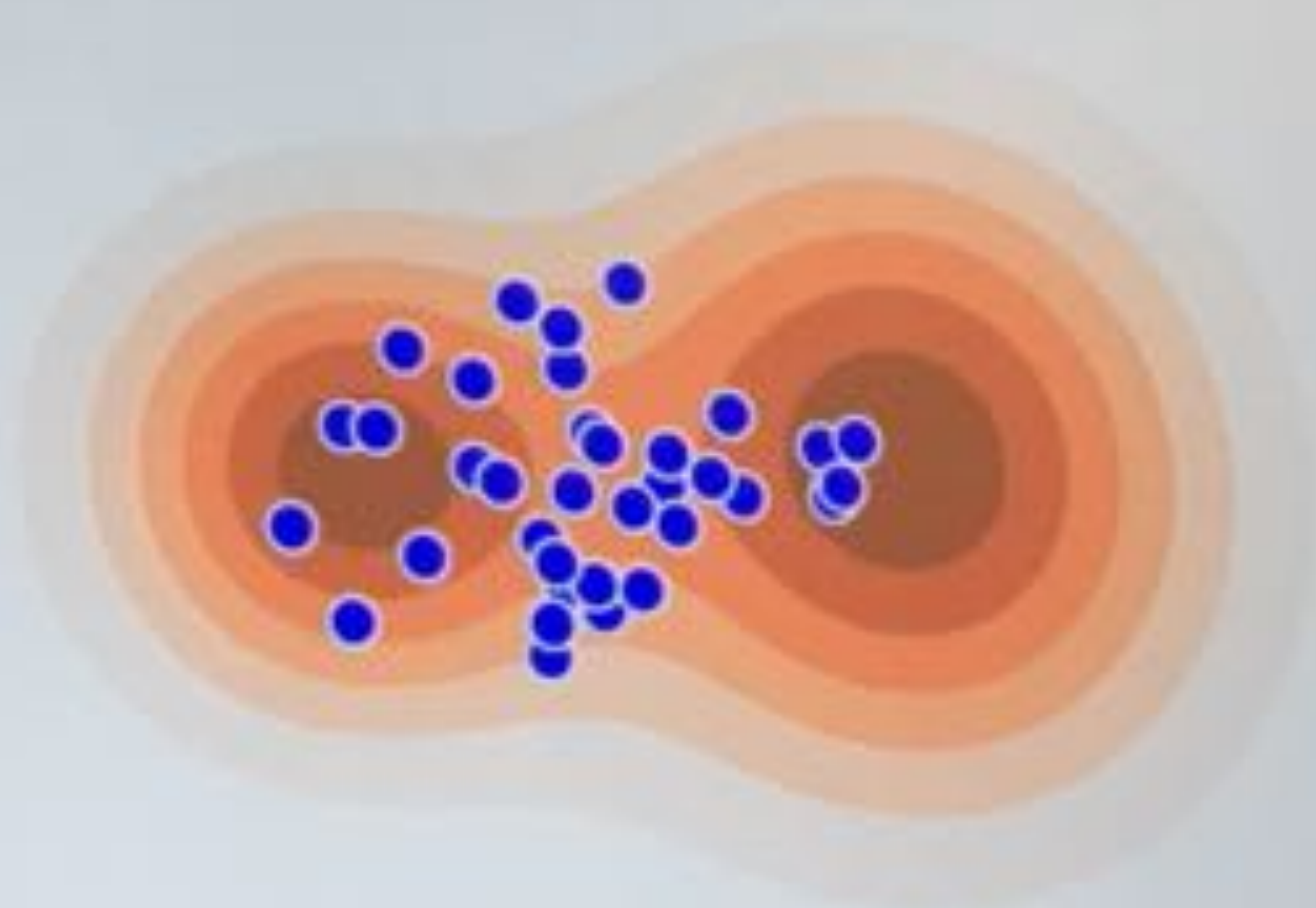
Problem: No sample from p . Cannot estimate $\mathbb{E}_{y \sim p}[k_v(y)]$.

$$(\text{Stein}) \text{ witness}(v) = \mathbb{E}_{x \sim q} [$$



Idea: Define T_p such that $\mathbb{E}_{y \sim p}(T_p k_v)(y) = 0$, for any v .

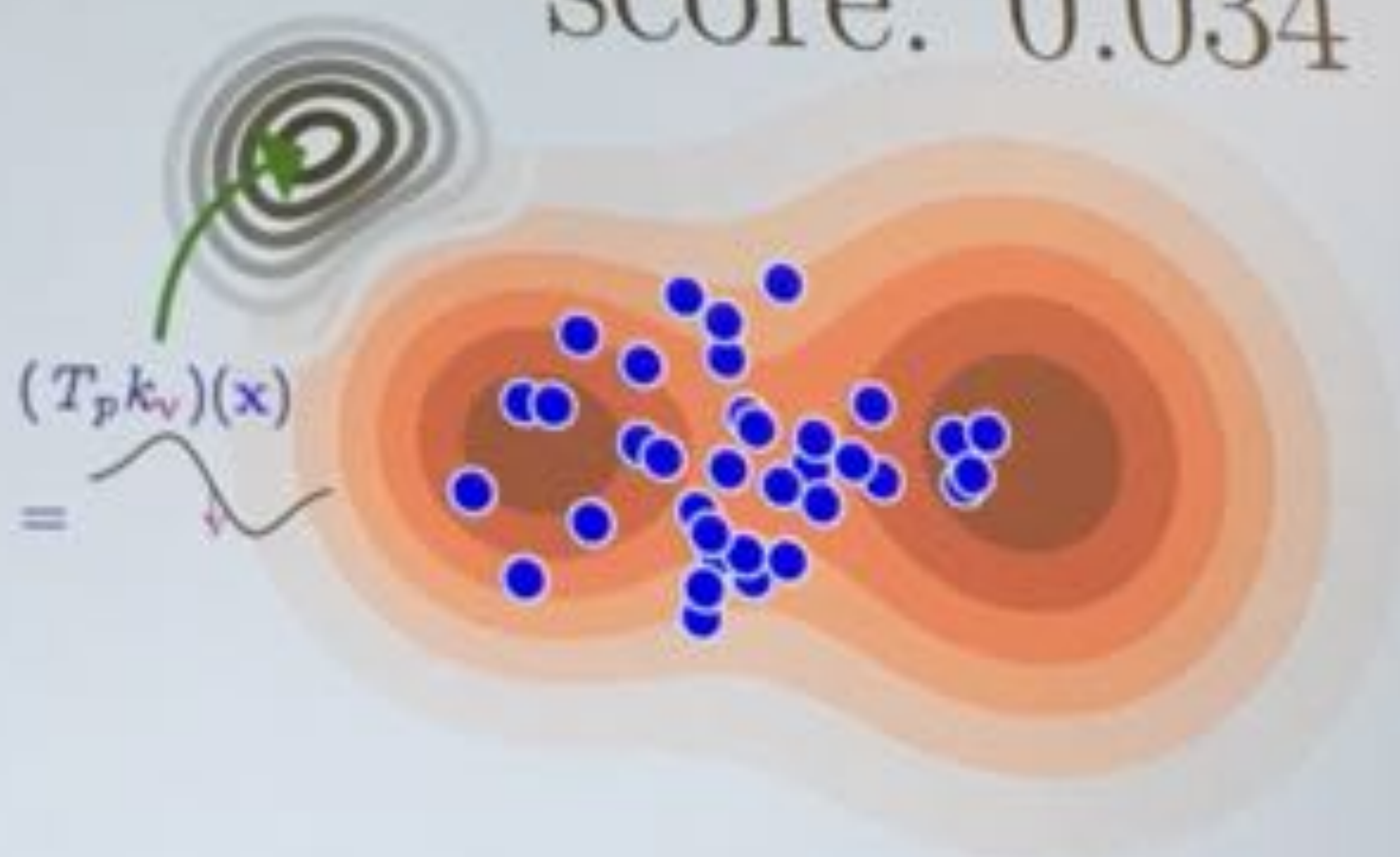
Proposal: Model Criticism with the Stein Witness



$$\text{score}(\mathbf{v}) = \frac{|\text{witness}(\mathbf{v})|}{\text{standard deviation}(\mathbf{v})}$$

Proposal: Model Criticism with the Stein Witness

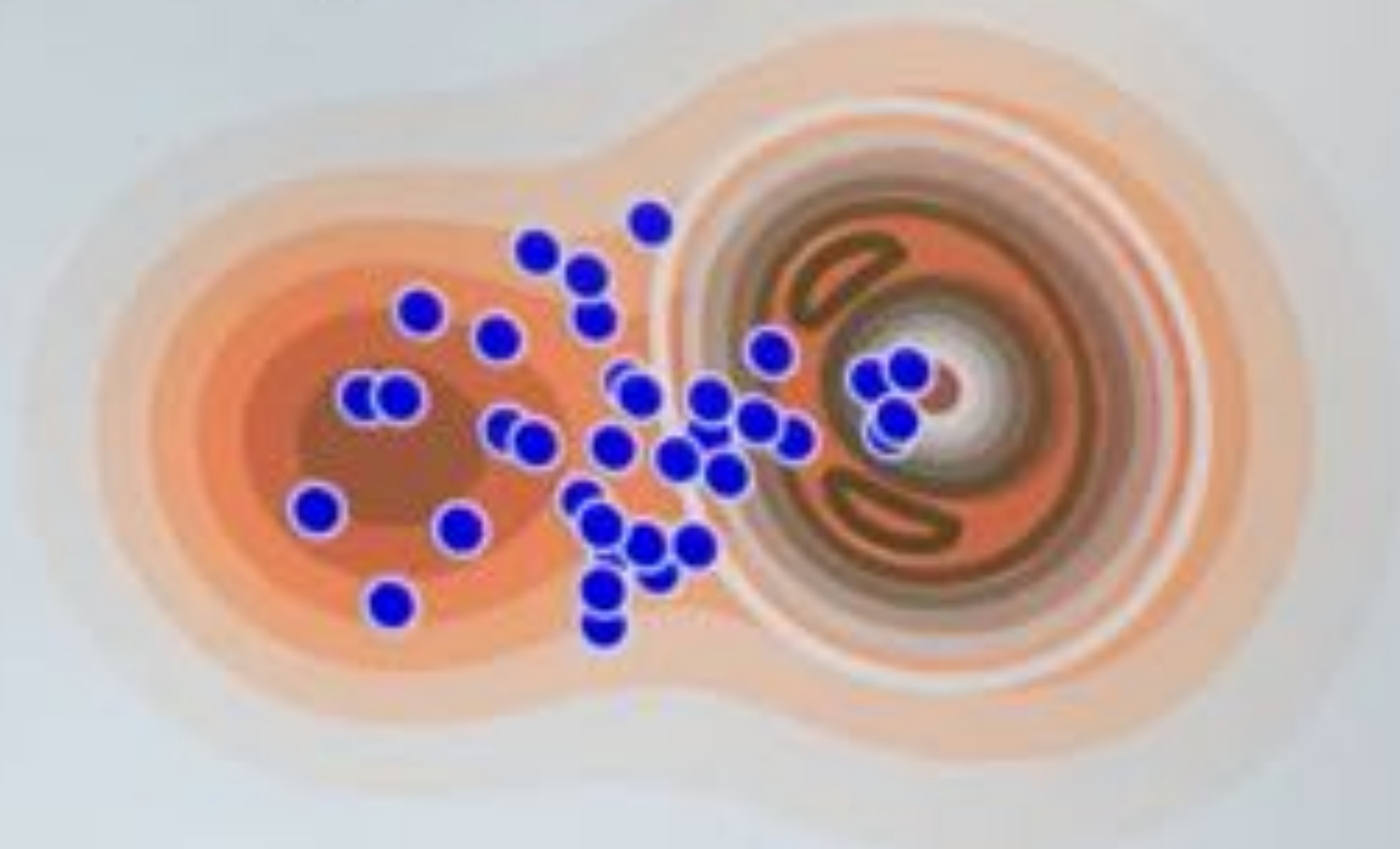
score: 0.034



$$\text{score}(\mathbf{v}) = \frac{|\text{witness}(\mathbf{v})|}{\text{standard deviation}(\mathbf{v})}$$

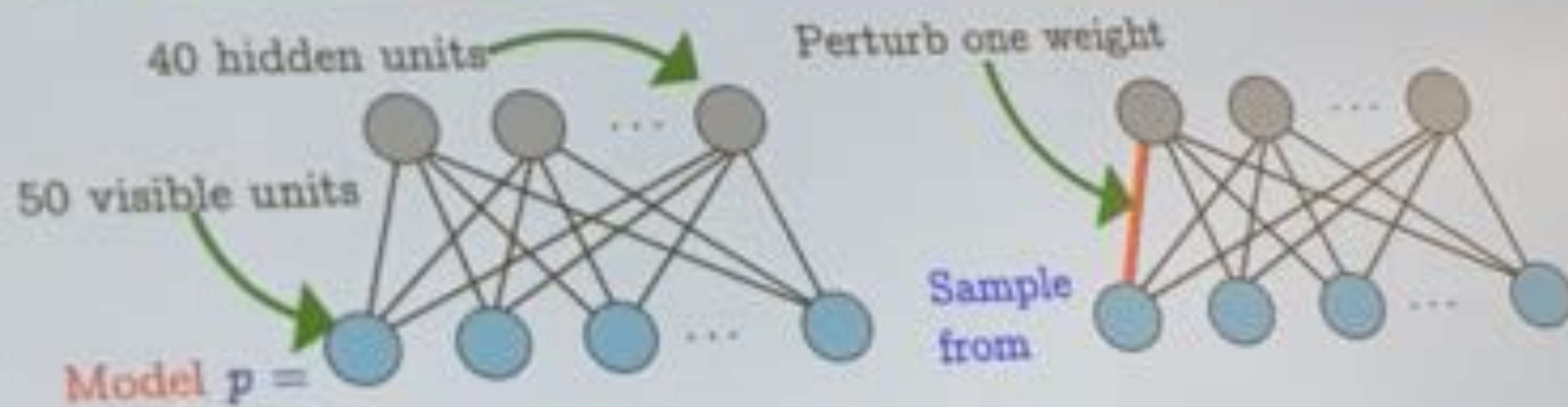
Proposal: Model Criticism with the Stein Witness

score: 0.44

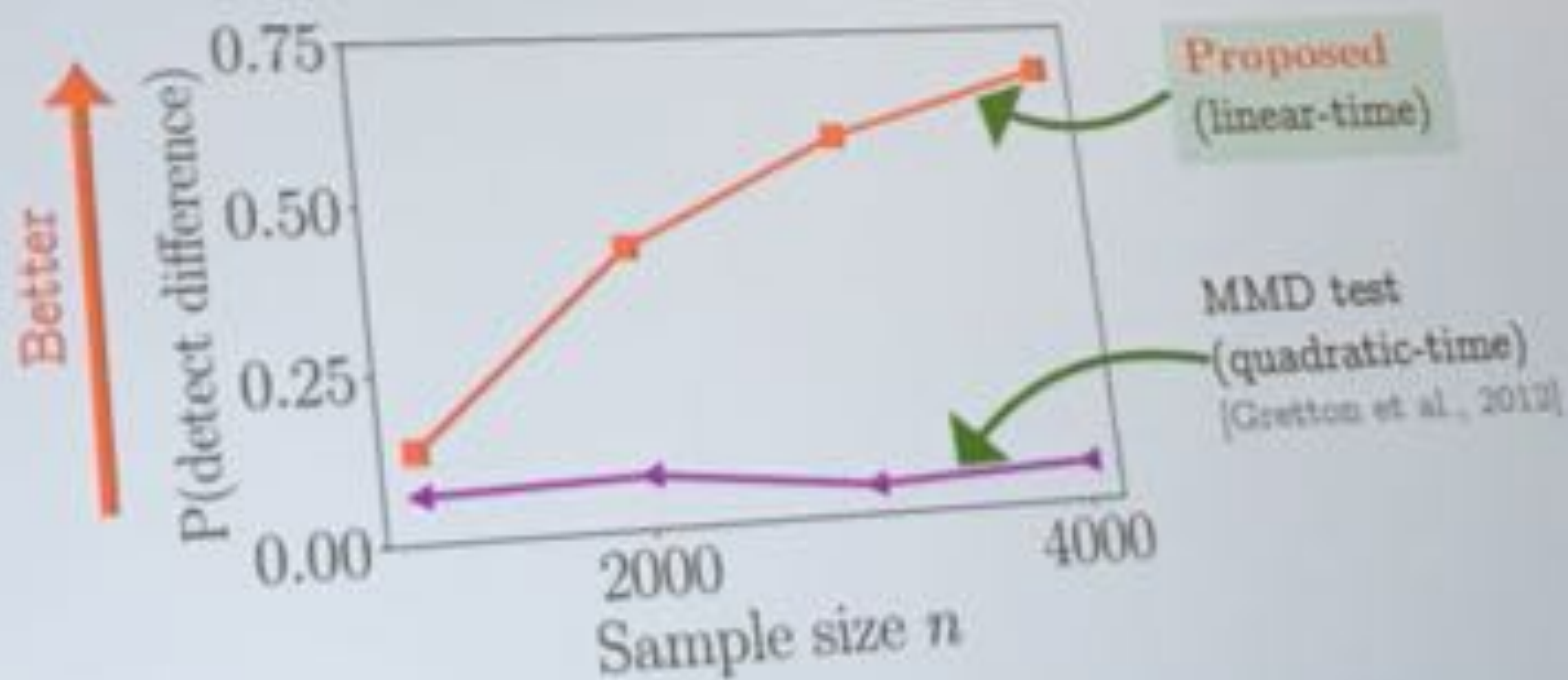
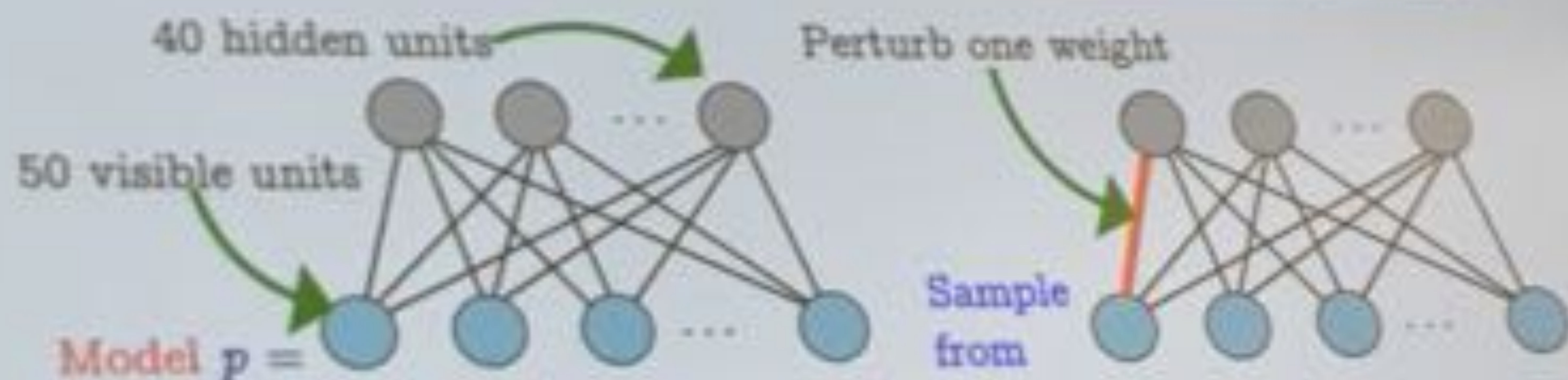


$$\text{score}(\mathbf{v}) = \frac{|\text{witness}(\mathbf{v})|}{\text{standard deviation}(\mathbf{v})}$$

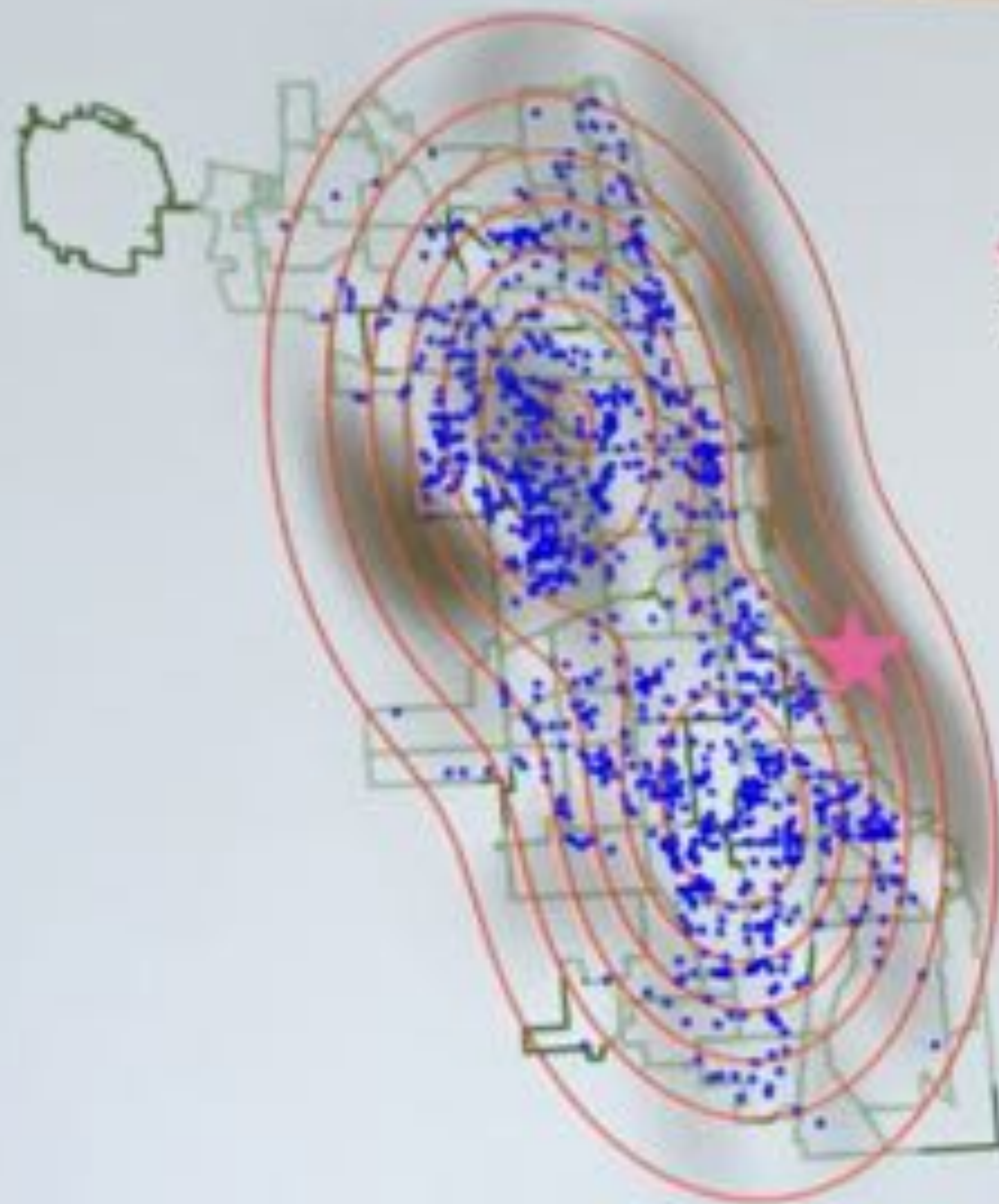
Experiment: Restricted Boltzmann Machine (RBM)



Experiment: Restricted Boltzmann Machine (RBM)



Interpretable Features: Chicago Crime

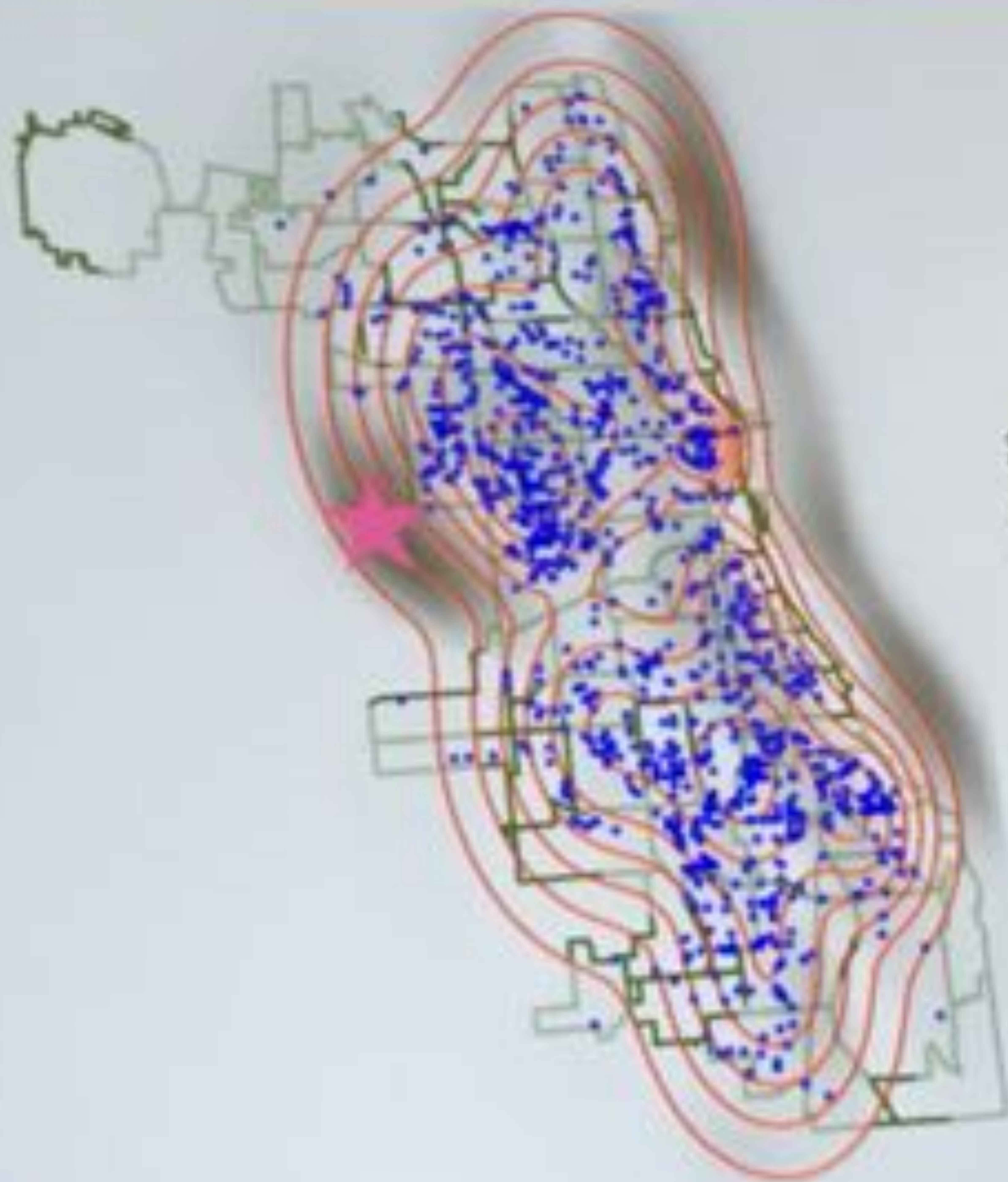


★ = optimized v .

No robbery in Lake Michigan.



Interpretable Features: Chicago Crime



Still, does not capture the left tail.

Conclusions

Proposed a new goodness-of-fit test.

- 1 Nonparametric. Normalizer not needed.
- 2 Linear-time
- 3 Interpretable

Poster #57 tonight

Python code: <https://github.com/wittawatj/kernel-gof>

