Machine Learning for High Energy Physics

Andrea De Simone

andrea.desimone@sissa.it

> The APP group at SISSA

interdisciplinary group working at interface of
particle physics, astrophysics and cosmology

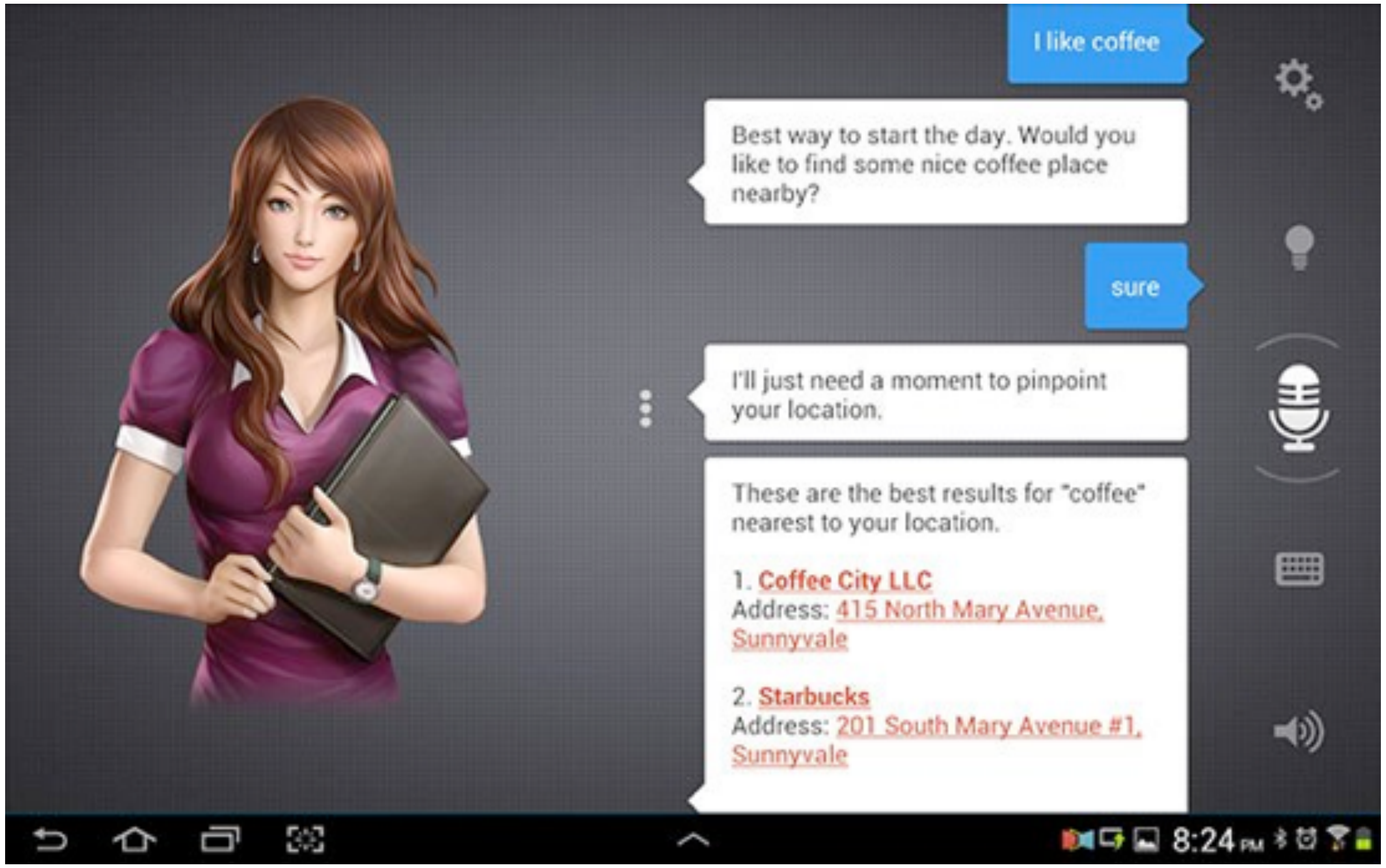
Address **fundamental issues** about our Universe:
origin and evolution, nature of gravity,
properties of dark matter and dark energy.

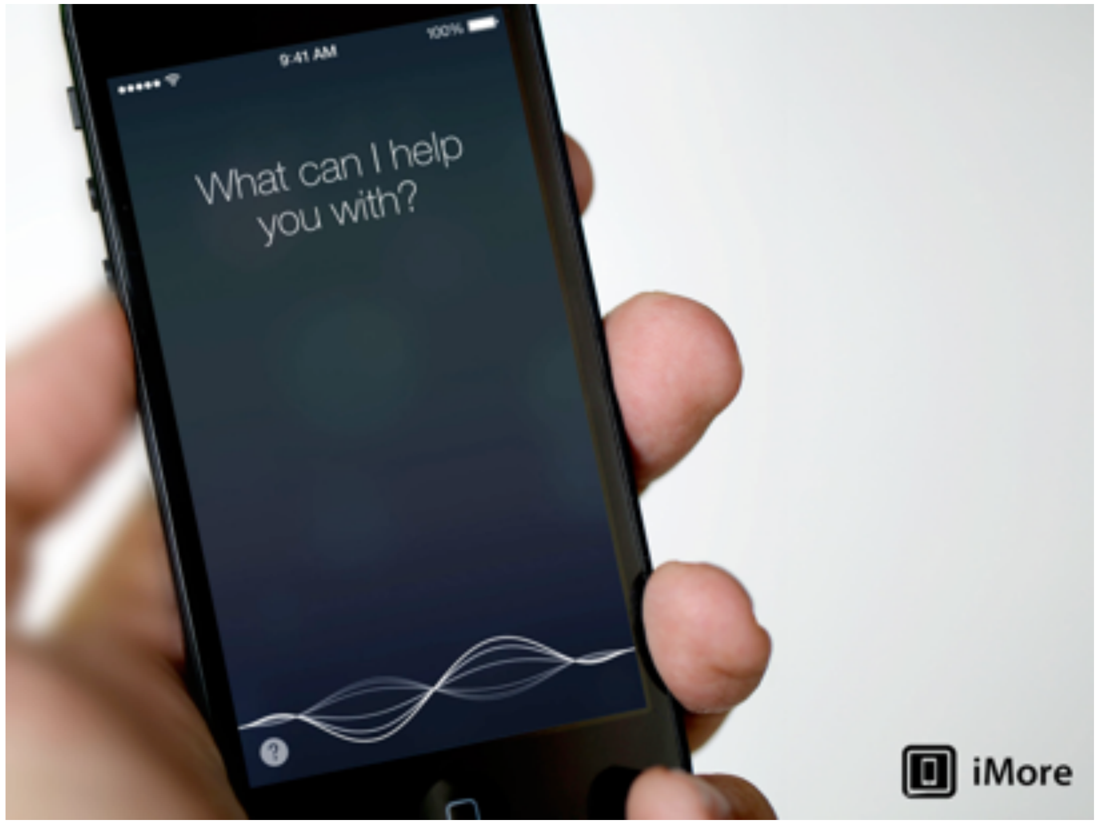
PIs in the group:
A. De Simone, S. Liberati,
P. Ullio, M. Viel

PIs affiliated from other SISSA groups:
C. Baccigalupi, R. Percacci, S. Petcov,
A. Romanino, P. Salucci

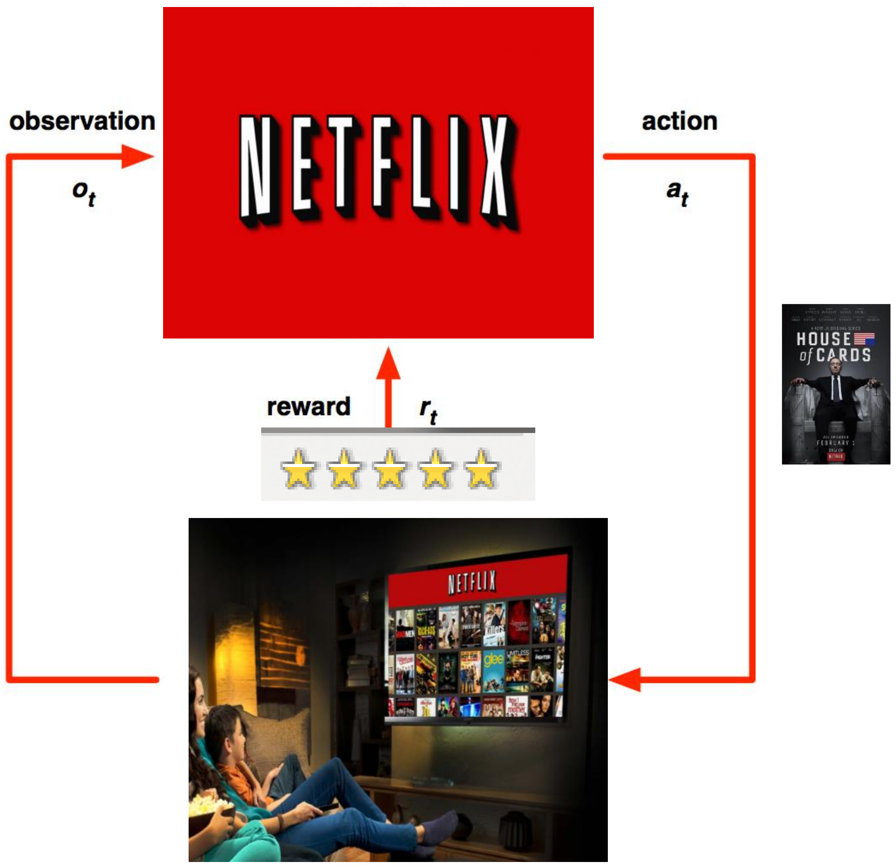
PIs affiliated form other institutions
P. Creminelli, E. Sefusatti

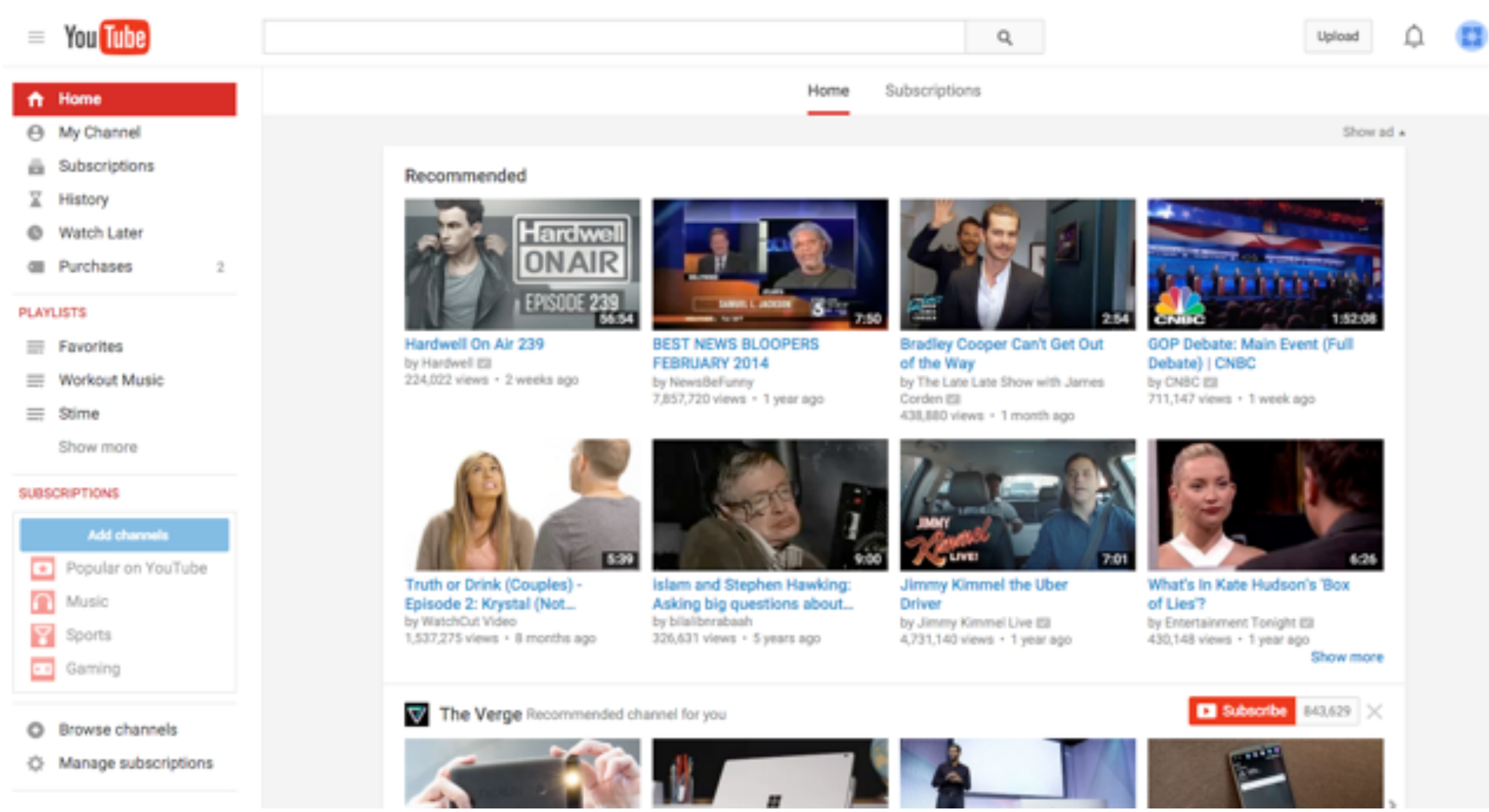
Speech recognition





Recommender systems





Creative Paintings

Which of these images were created by a machine?

[Elgammal et al - 1706.07068]

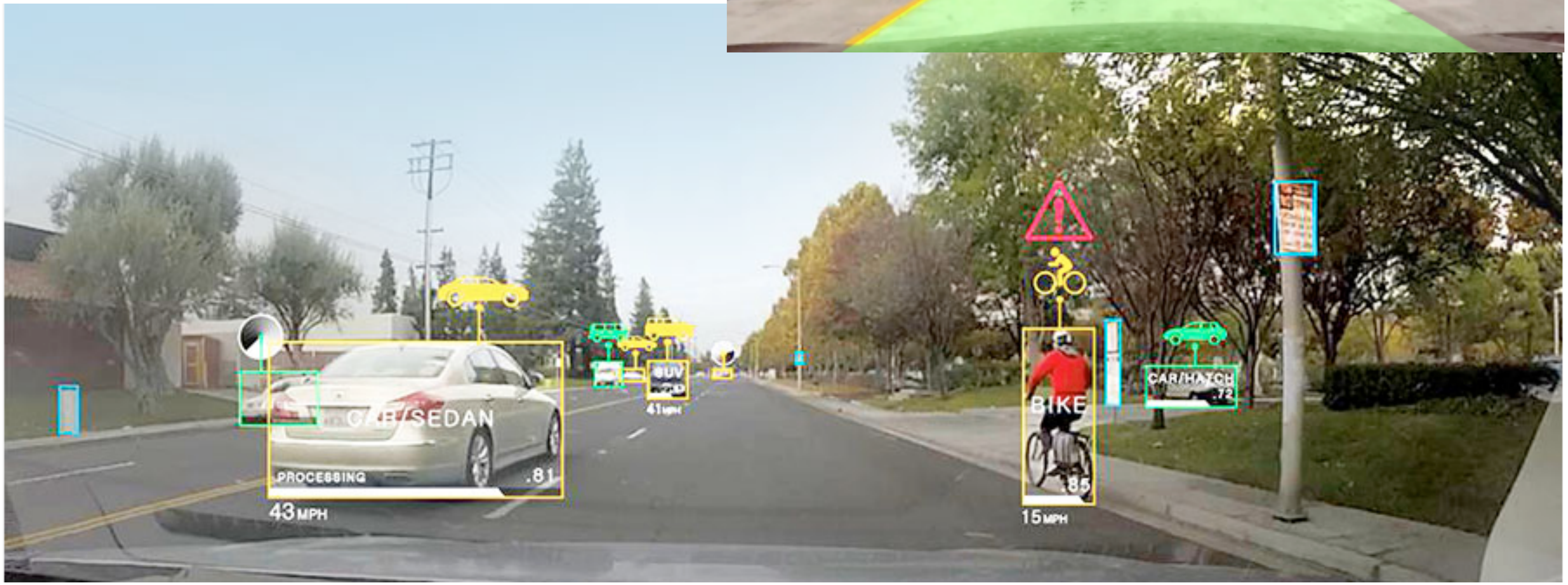
Generating Faces

These people do not exist!

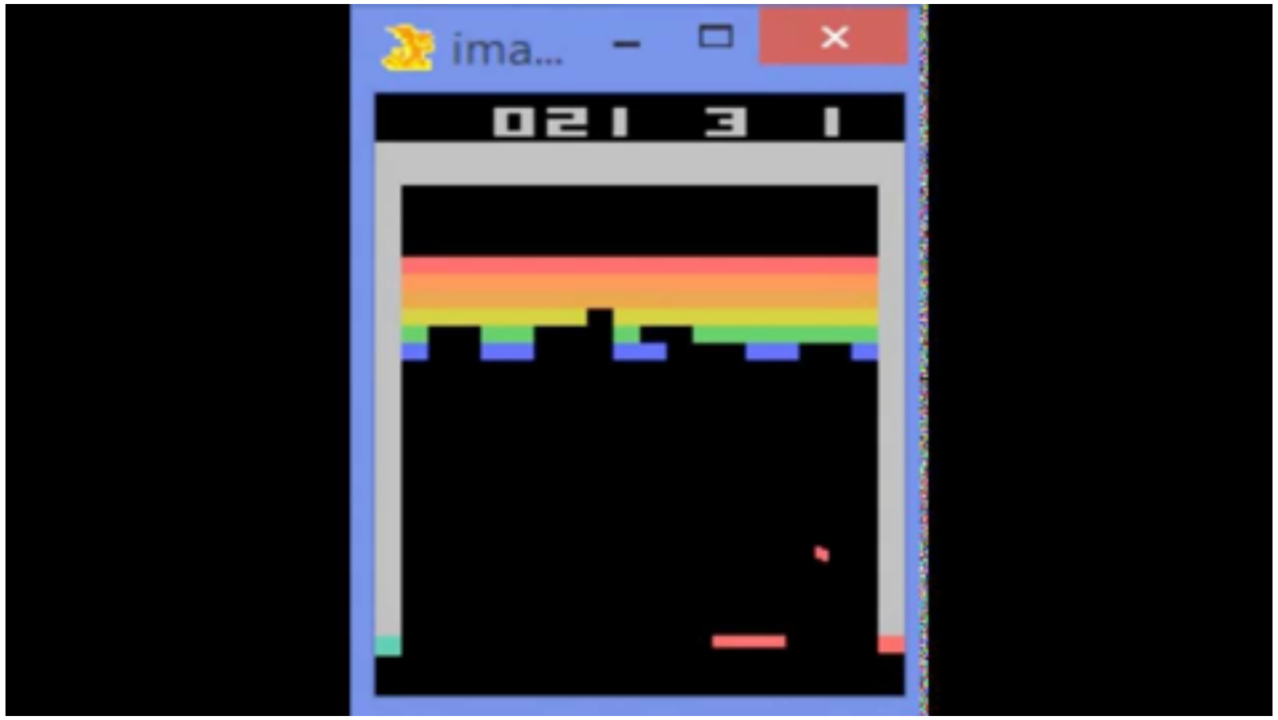
CelebA-HQ: https://youtu.be/XOxxPcy5Gr4

[Karras et al - 1710.10196]

Autonomous driving



Games



Google's DeepMind
plays Breakout

AlphaGo beats world champion Go 4-1

2010-05-06, 2:45pm

2010-05-06, 2:45pm

Flash Crash!

Dow-Jones Industrial Average

lost ~9% in 36 minutes!

> Outline

1. Machine Learning in Science

2. Open problems in High-Energy Physics (HEP)

3. Statistical test of dataset compatibility

4. Applications to HEP

> Outline

1. Machine Learning in Science

2. Open problems in High-Energy Physics (HEP)

3. Statistical test of dataset compatibility

4. Applications to HEP

> Machine Learning

“A computer program is said to *learn* from **experience E**
with respect to some class of **tasks T**
and **performance measure P**,

if its performance at tasks in T, as measured by P,
improves with experience E.”

[Mitchell - 1997]

Why is Machine Learning so cool?

- a guide through big data (data mining)
- many diverse applications
  (from engineering to commerce to science)
- can help making our life better/easier
- …
- **can help scientists to do science better and faster**

> Why now?

- More powerful machines

both speed and storage

- More data

almost everything is recorded!

- Easier access

internet revolution, easier to share big data

> Machine Learning

Supervised Learning

Unsupervised Learning

Reinforcement Learning

> Supervised Learning

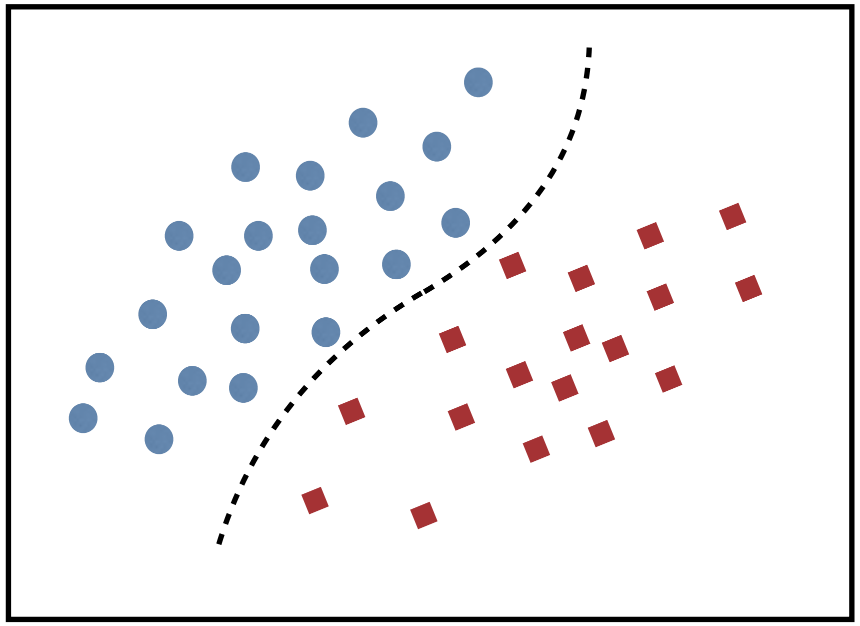
labelled data

$$\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$$

features

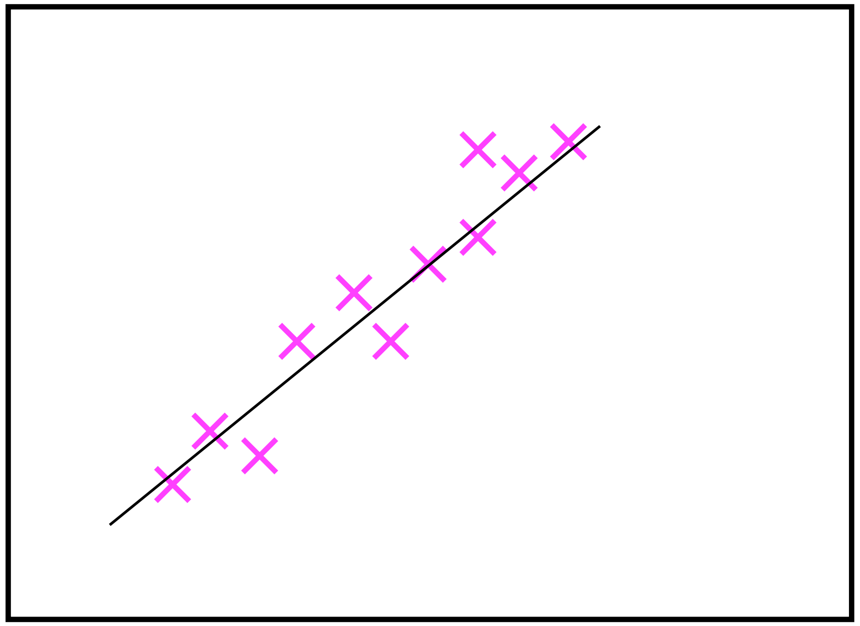
labels

Classification



Logistic Regression
Neural Networks
Decision Trees
Nearest Neighbors
…

Regression



Polynomial Regression
Neural Networks
Support Vector Machines
Nearest Neighbors
…

machine “learns” the model

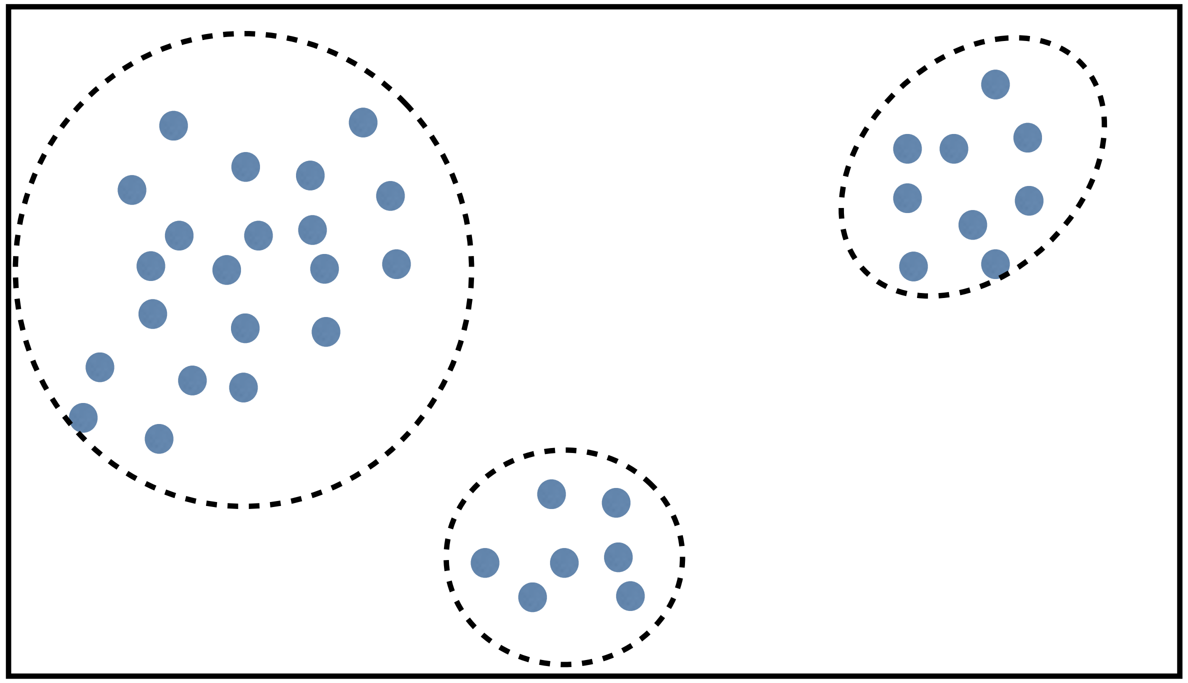
$$f(x) = y$$

> Unsupervised Learning

unlabelled data

$\{\boldsymbol{x}_i\}_{i=1}^N$

features



Cluster Analysis
Dimensionality Reduction
Anomaly Detection
…

machine “learns”
patterns, structures, representations, etc.
of the data

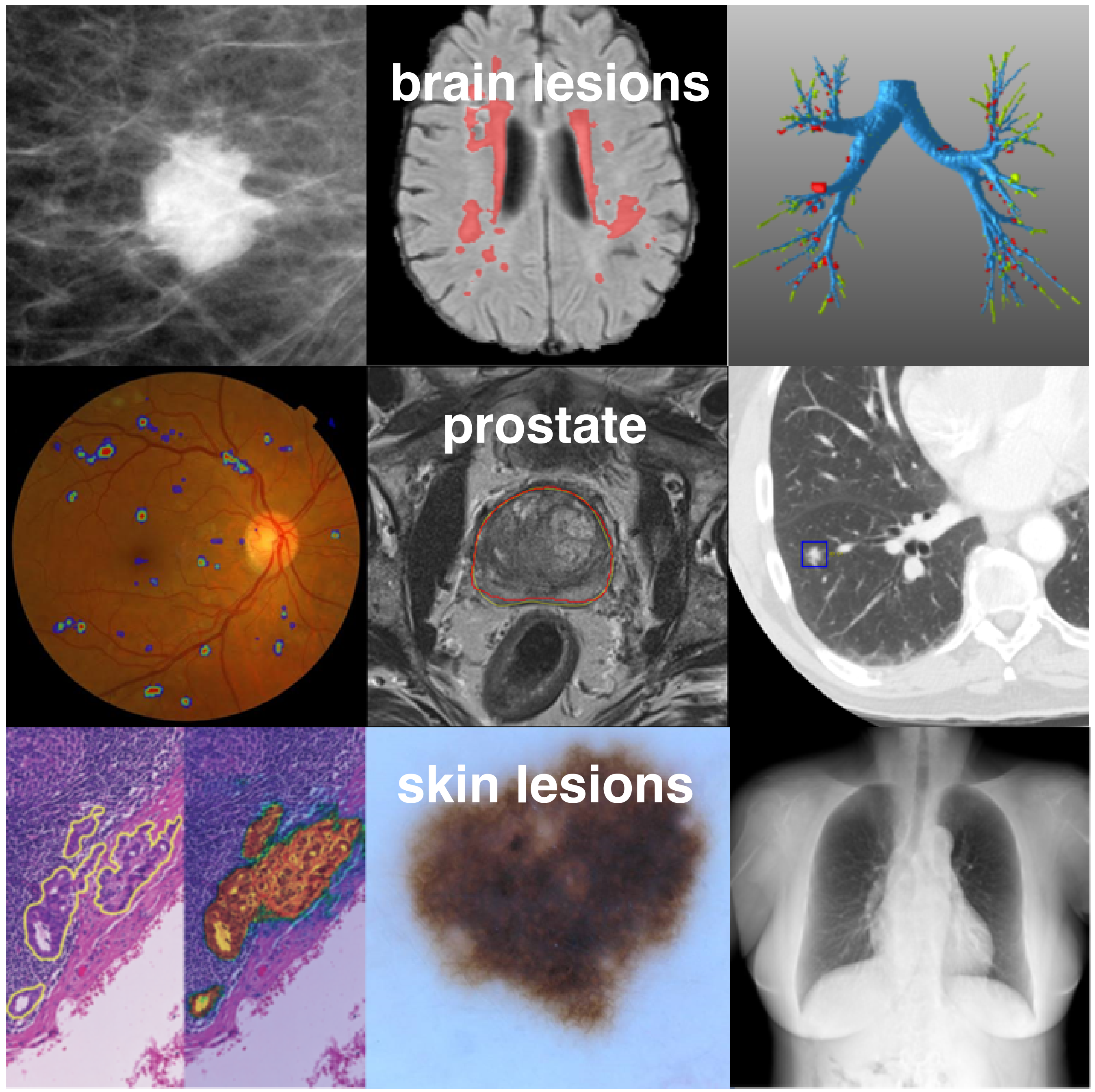
> ML in Medical Science

deep learning
achieved state-of-the art results

mammography

diabetic retinopaty

breast cancer metastases



airways

lung nodule

bone suppression

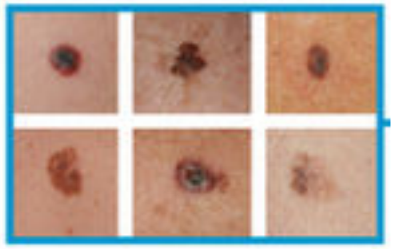
[Litjens et al. - Medical Image Analysis 2017]

Dermatologist-level classification of skin cancer with deep neural networks

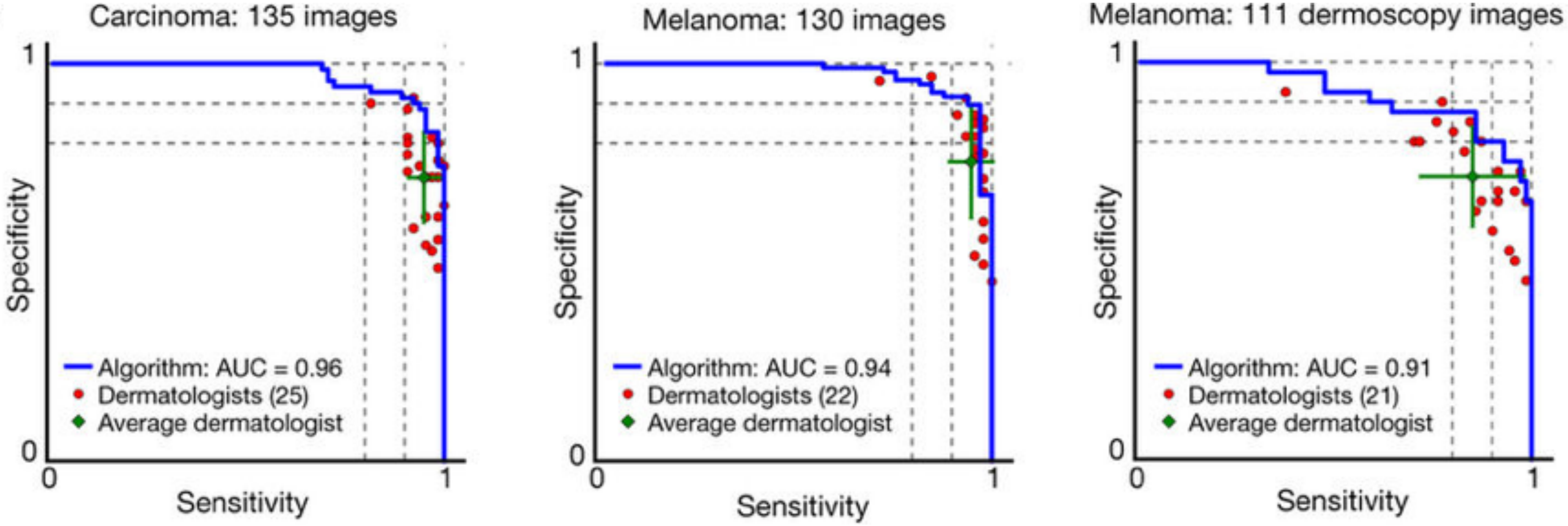
$$\textbf{sensitivity} = \frac{\text{True Positives}}{\text{Positives}}$$

$$\textbf{specificity} = \frac{\text{True Negatives}}{\text{Negatives}}$$

Melanomas

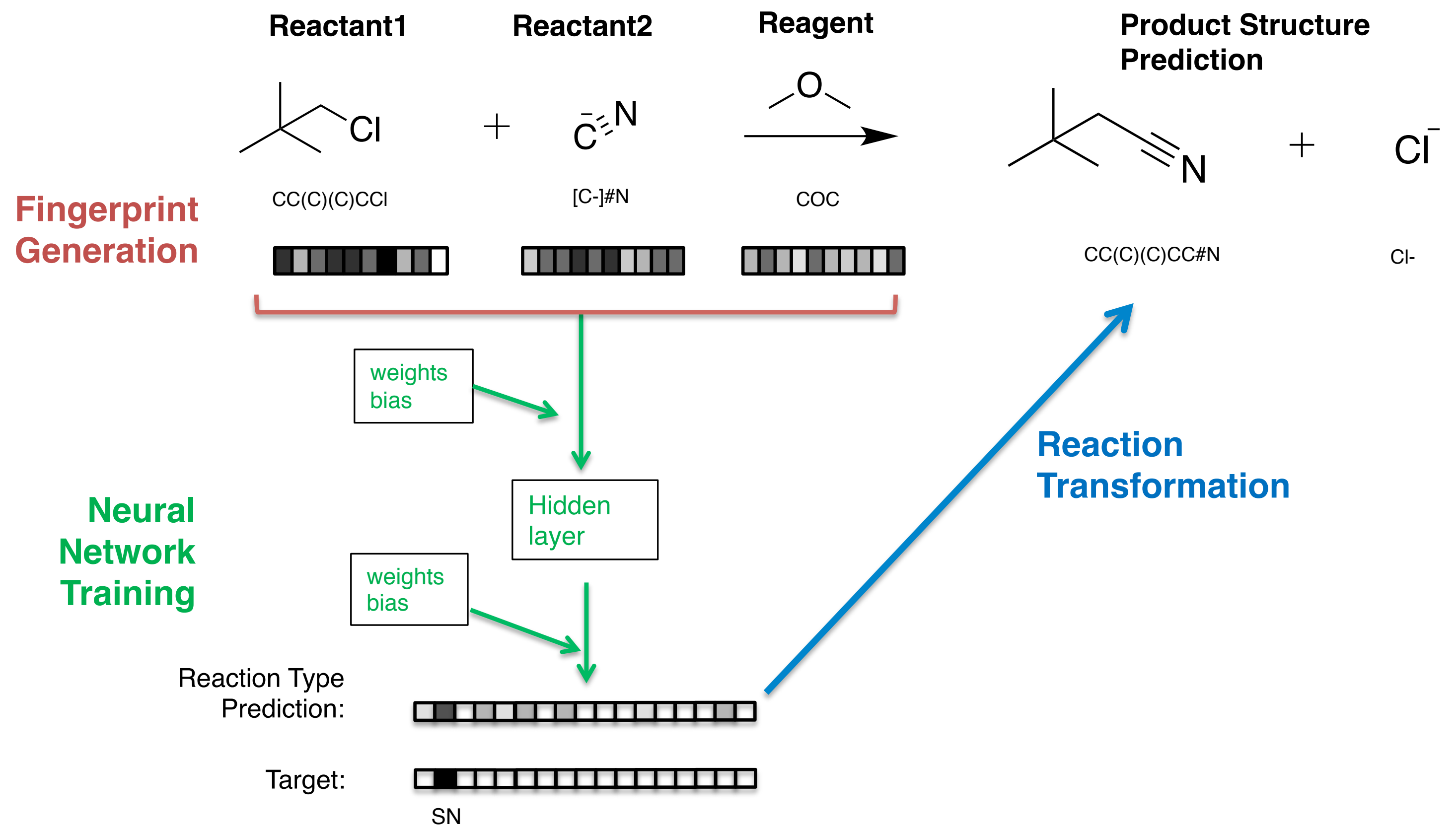


Basal cell carcinomas



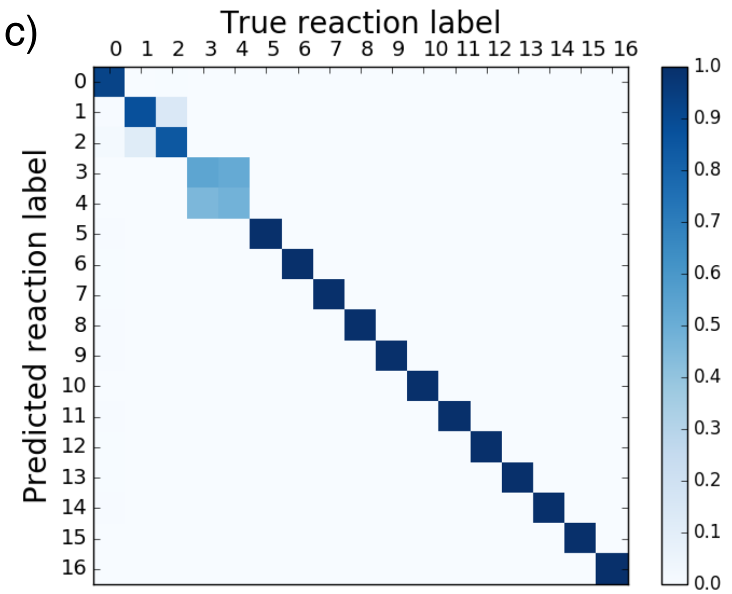
[Esteva et al. - Nature 2017]

Neural Networks for the prediction of organic chemistry reactions



predict probability of 17 different reaction types

0. Null Reaction
1. Nucleophilic substitution
2. Elimination
3. Nucleophilic Substitution with Methyl Shift
4. Elimination with methyl shift
5. Hydrohalogenation (Markovnikov)
6. Hydrohalogenation (Anti-Markovnikov)
7. Hydration (Markovnikov)
8. Hydration (Anti-Markovnikov)
9. Alkoxymercuration-demercuration
10. Hydrogenation
11. Halogenation
12. Halohydrin formation
13. Epoxidation
14. Hydroxylation
15. Ozonolysis
16. Polymerization

c)


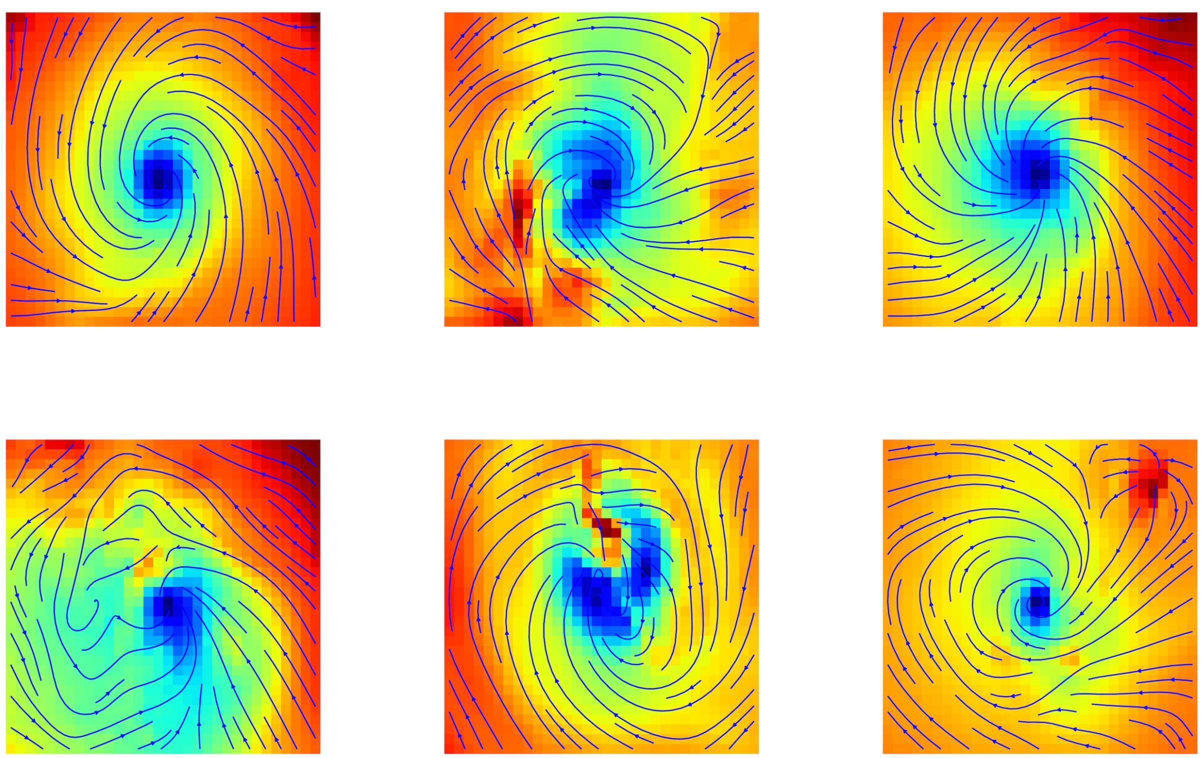
(~85% accuracy on test set)

[Wei, Duvenaud, Aspuru-Guzik, 2016]

Detect extreme weather
using deep learning

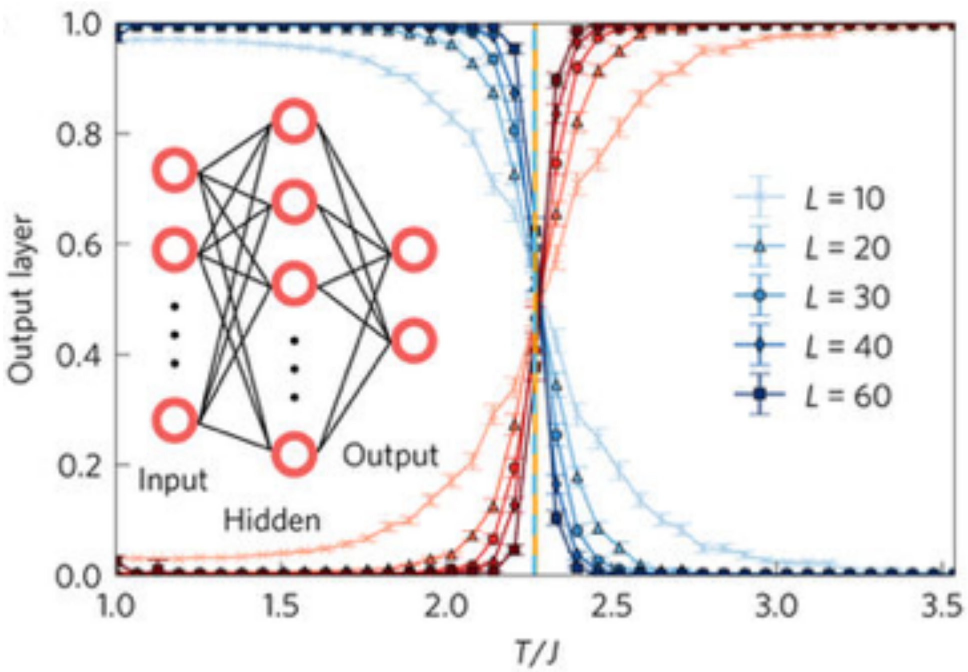
classification accuracy

Event Type	Train	Test	Train time
Tropical Cyclone	99%	99%	$\approx$ 30 min
Atmospheric River	90.5%	90%	6-7 hour
Weather Front	88.7%	89.4%	$\approx$ 30 min



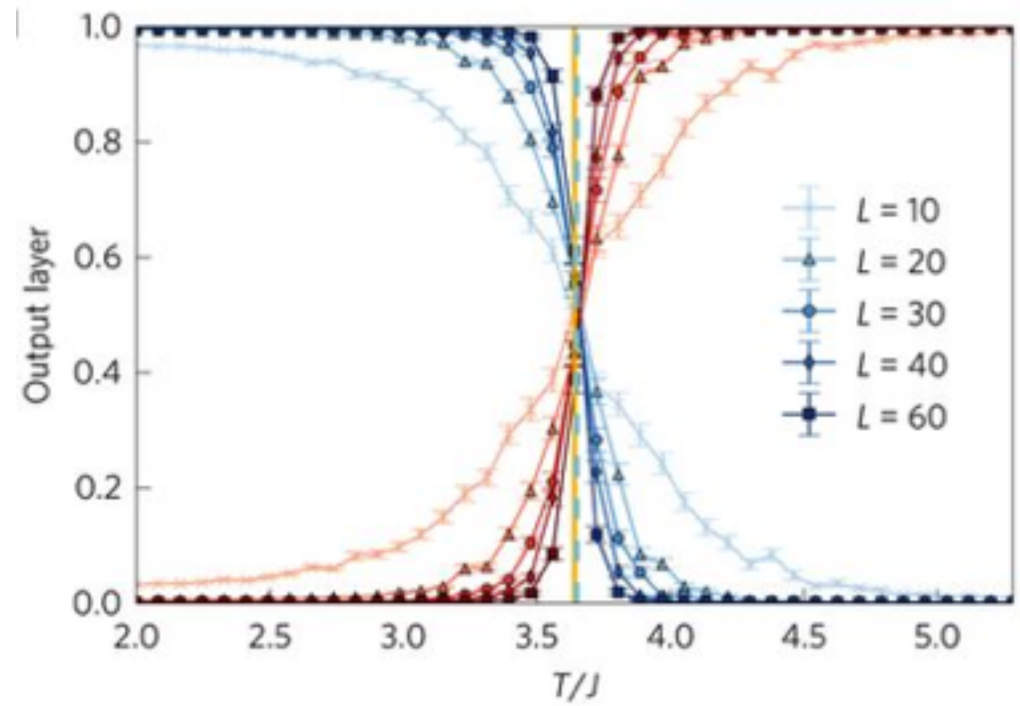
[Liu et al - 1605.01156]

Machine learning phases of matter

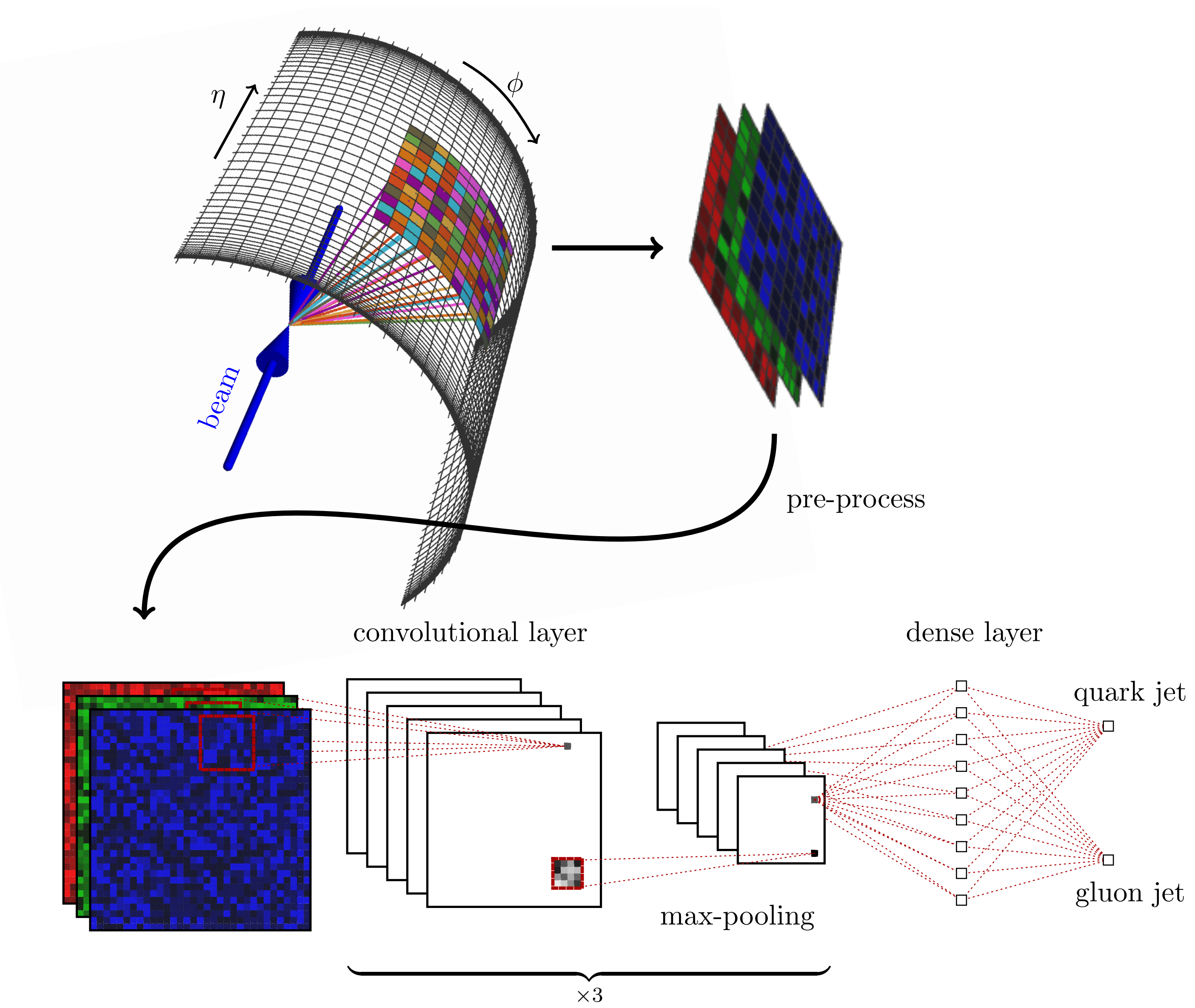


square-lattice Ising model:
Tc/J = 2.266 ± 0.002
(exact: 2.269..)

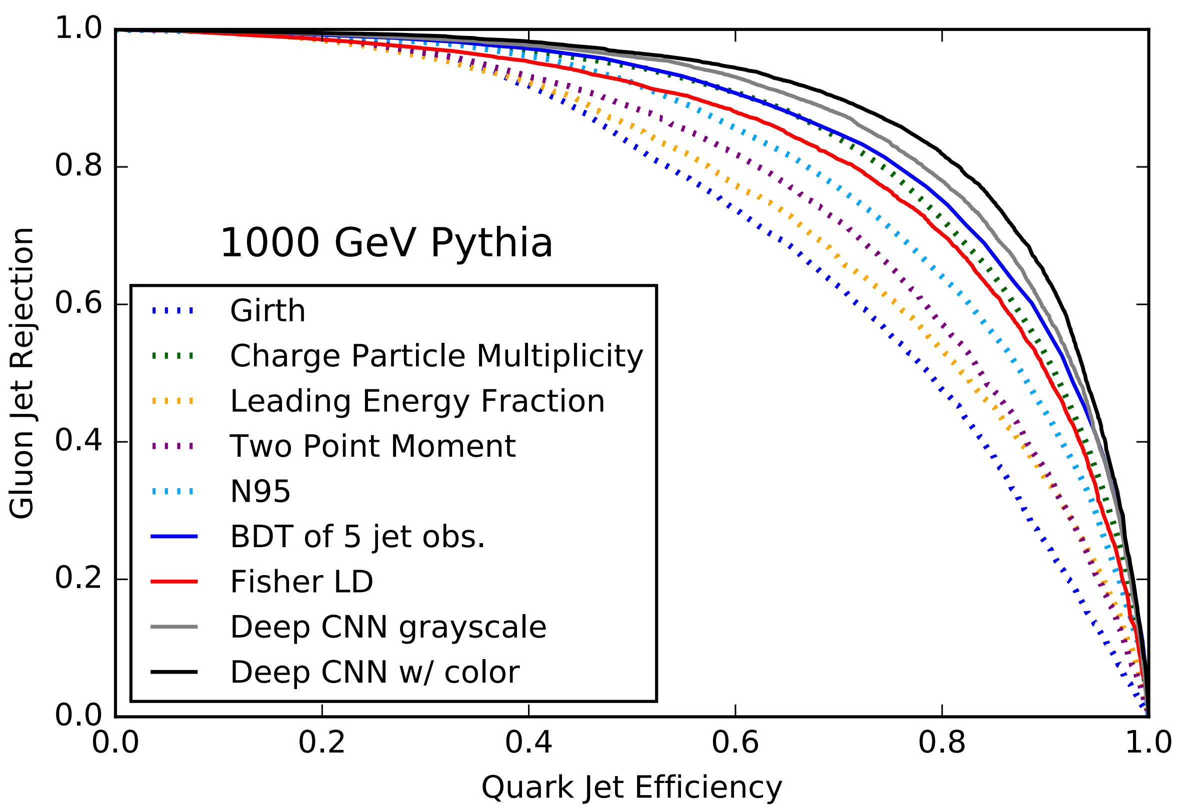
triangular-lattice Ising model:
Tc/J = 3.65 ± 0.01
(exact: 3.64095…)



[Carrasquilla, Melko - Nature 2017]



quark/gluon jet discrimination



[Komiske, Metodiev, Schwartz - 1612.01551]

> Outline

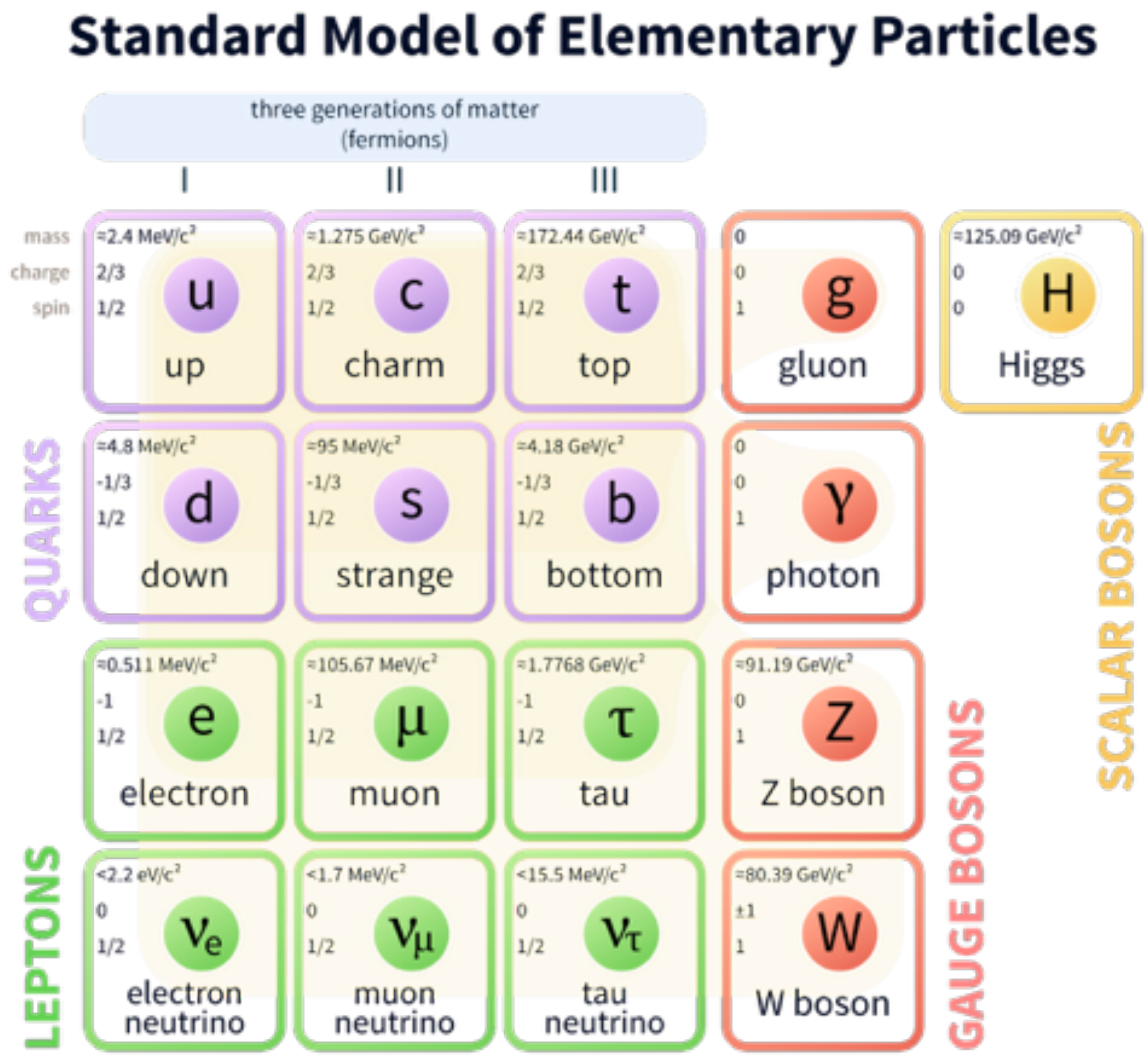
1. Machine Learning in Science

2. Open problems in High-Energy Physics (HEP)

3. Statistical test of dataset compatibility

4. Applications to HEP

> The Standard Model



Excellent description of sub-nuclear phenomena

One of the most accurately verified predictions in history of physics

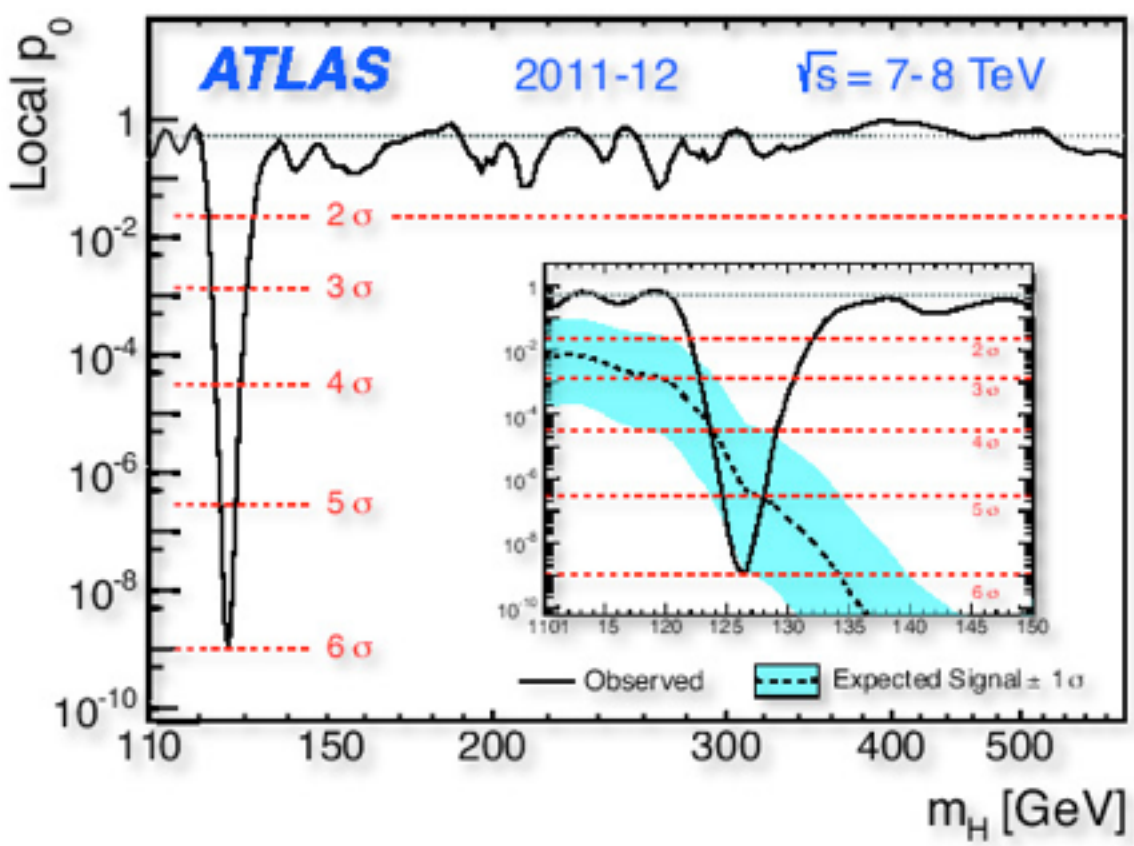
anomalous magnetic moment of the electron

$$a_{exp} = 0.001\ 159\ 652\ 180\ 91(26)$$
$$a_{th} = 0.001\ 159\ 652\ 181\ 643(764)$$

more than 10 significant figures!

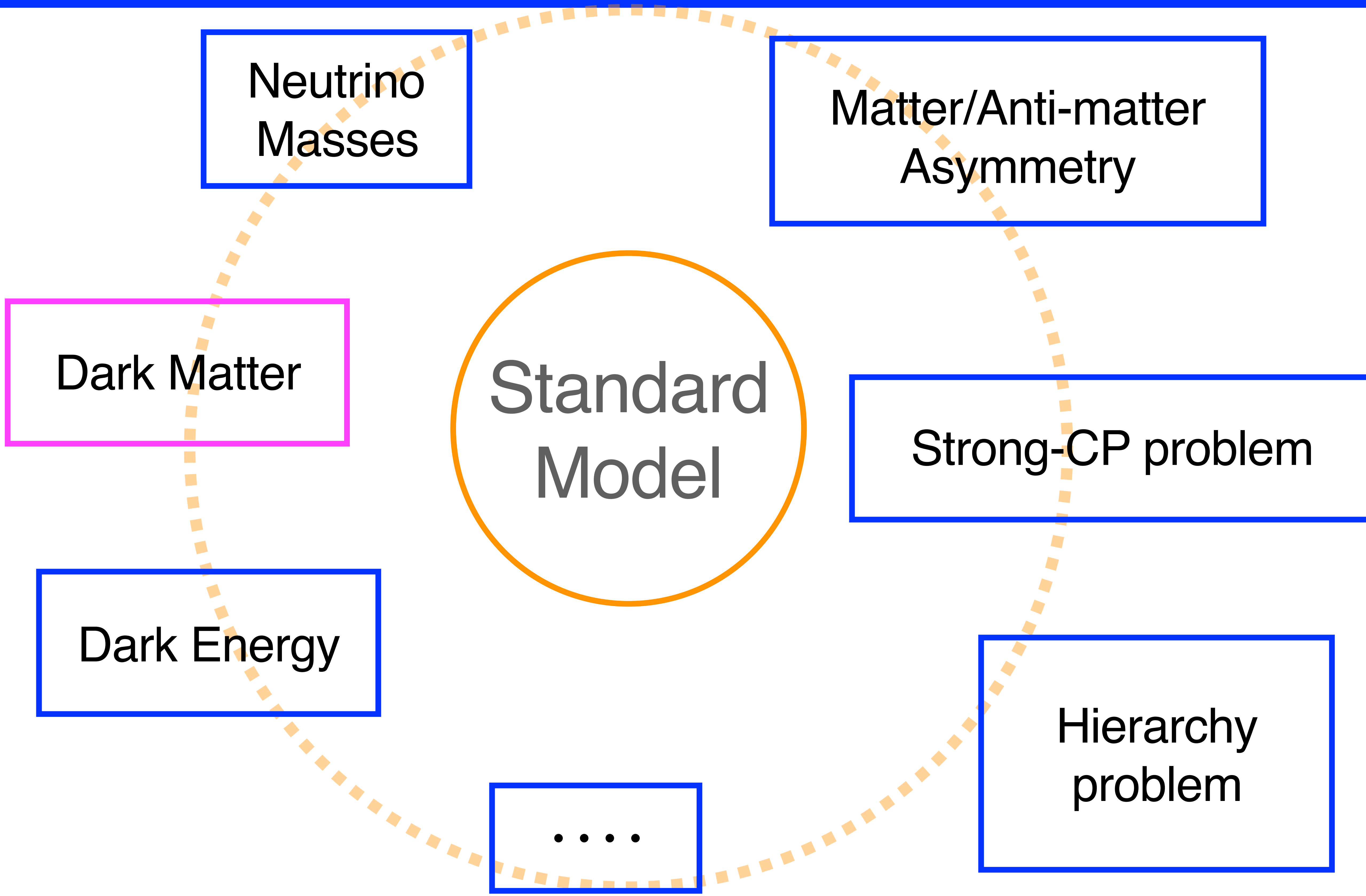
> Higgs Discovery

2012-07-04



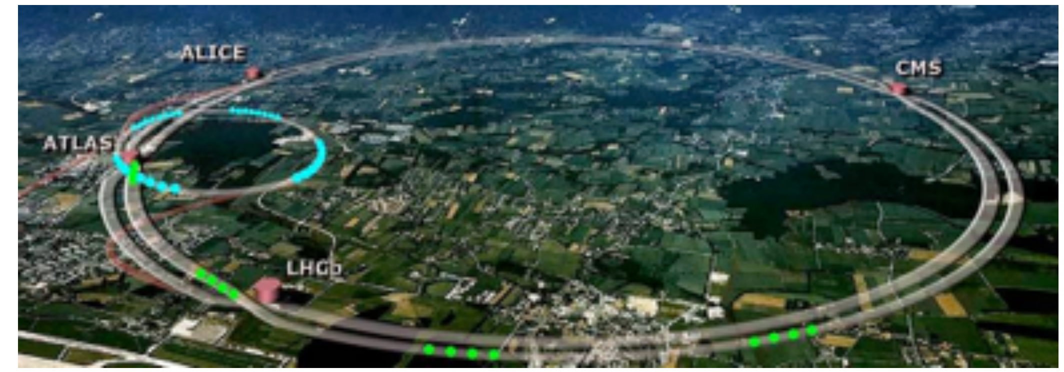
Is there anything left to discover?

> Need to go beyond...



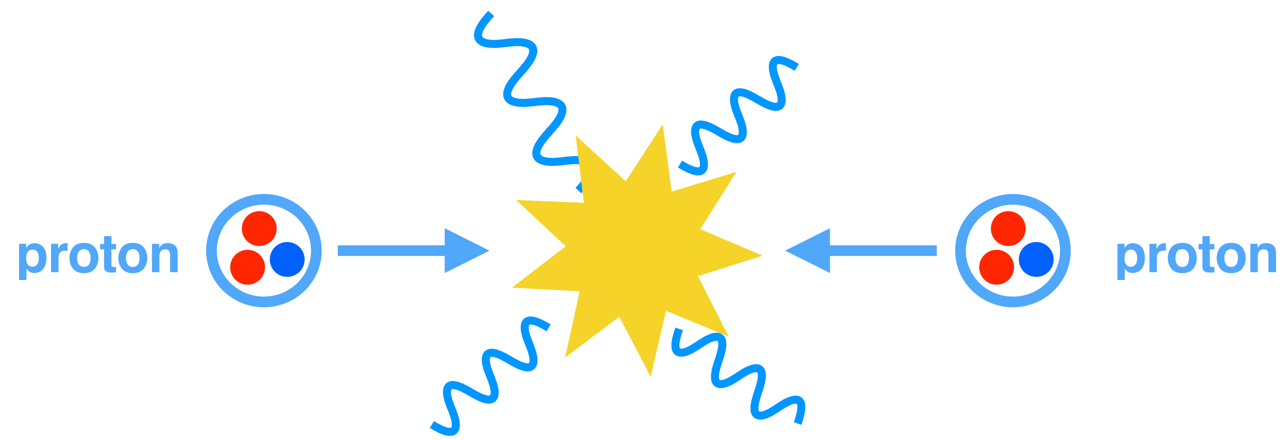
The most complex (and expensive) experiment ever built!

Cost ~ 4 GigaEur



counter-rotating proton beams in 27km circumference ring

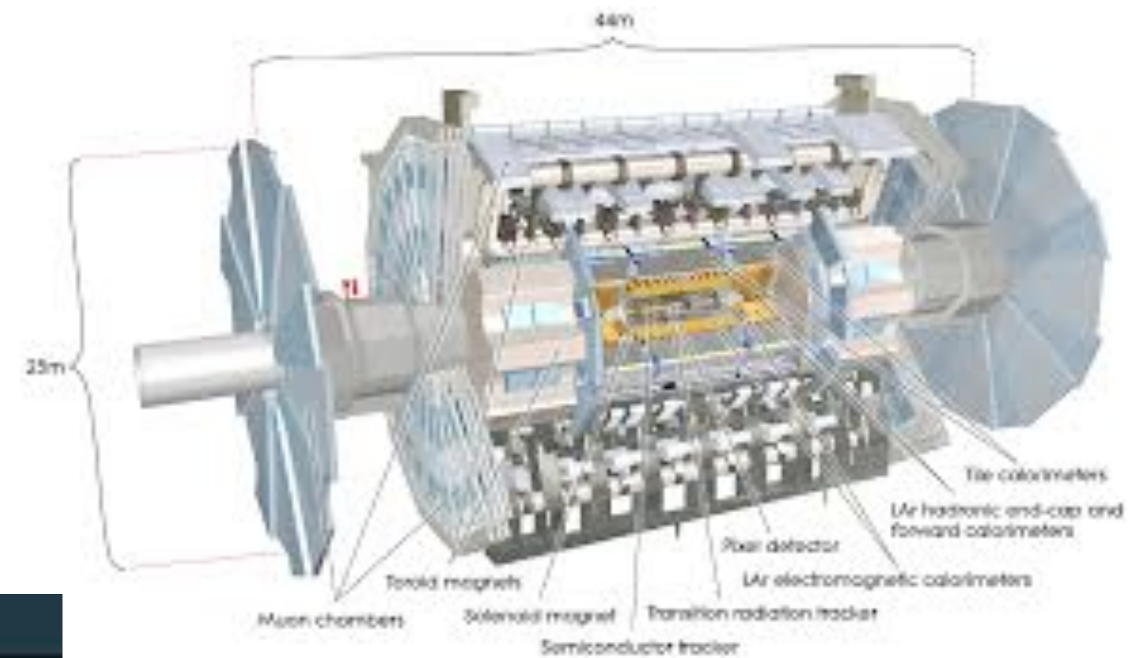
center-of-mass energy: 13 TeV

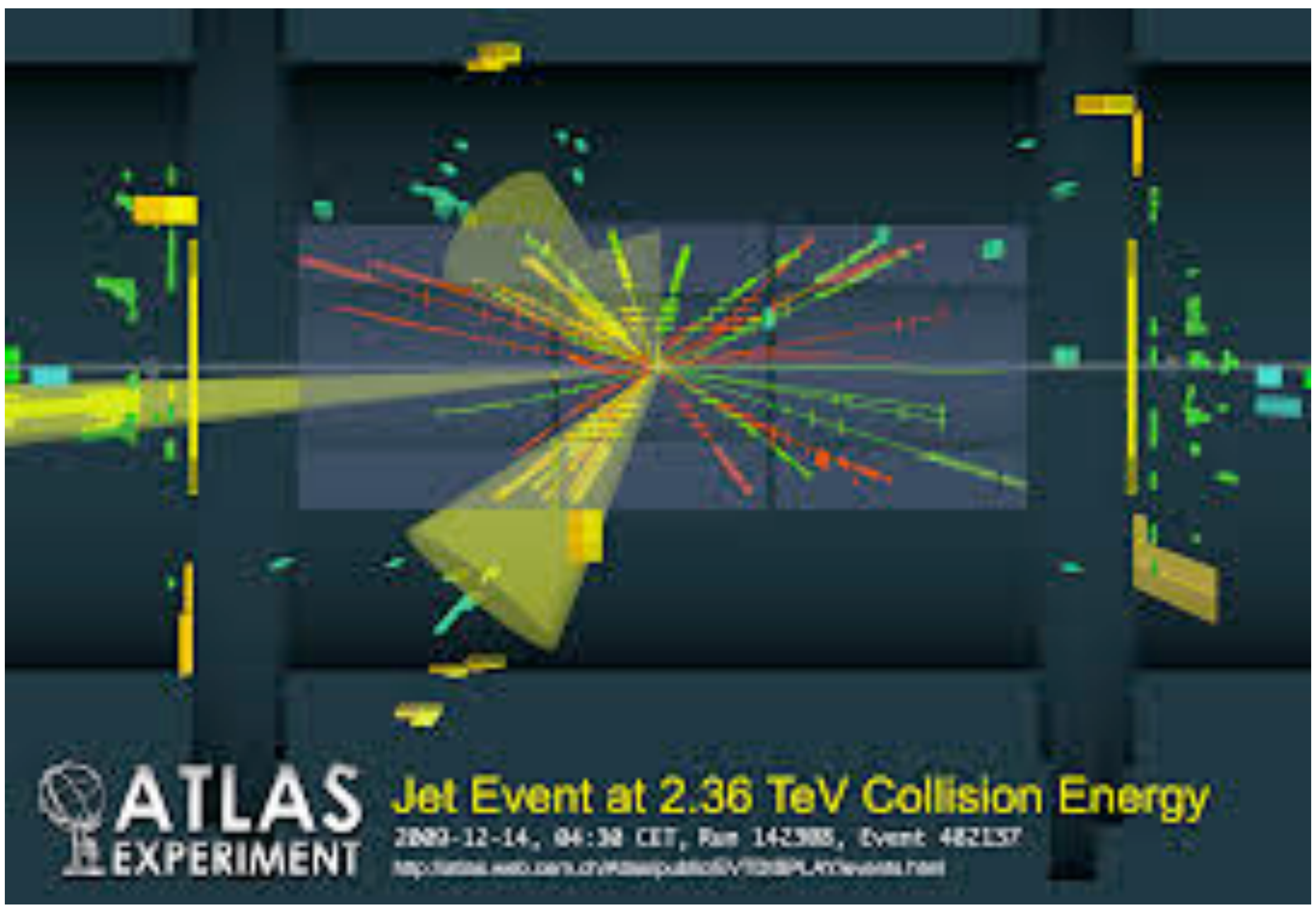


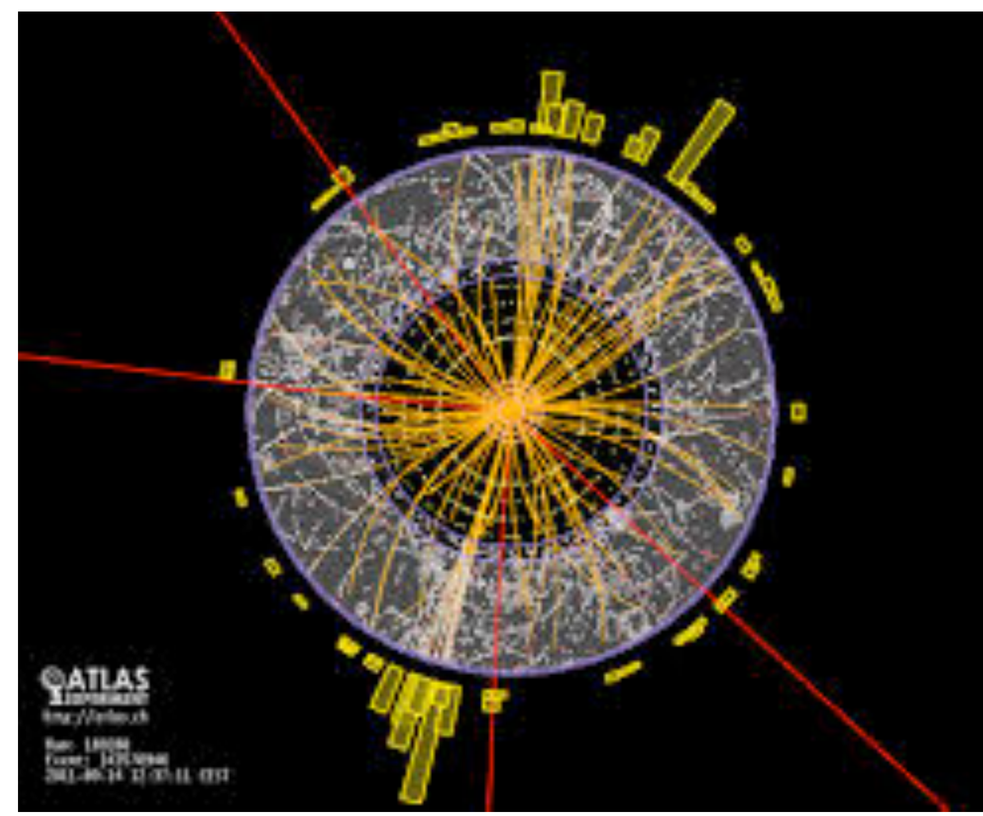
detectors at 4 collision points:

- ATLAS
- CMS
- LHCb
- ALICE

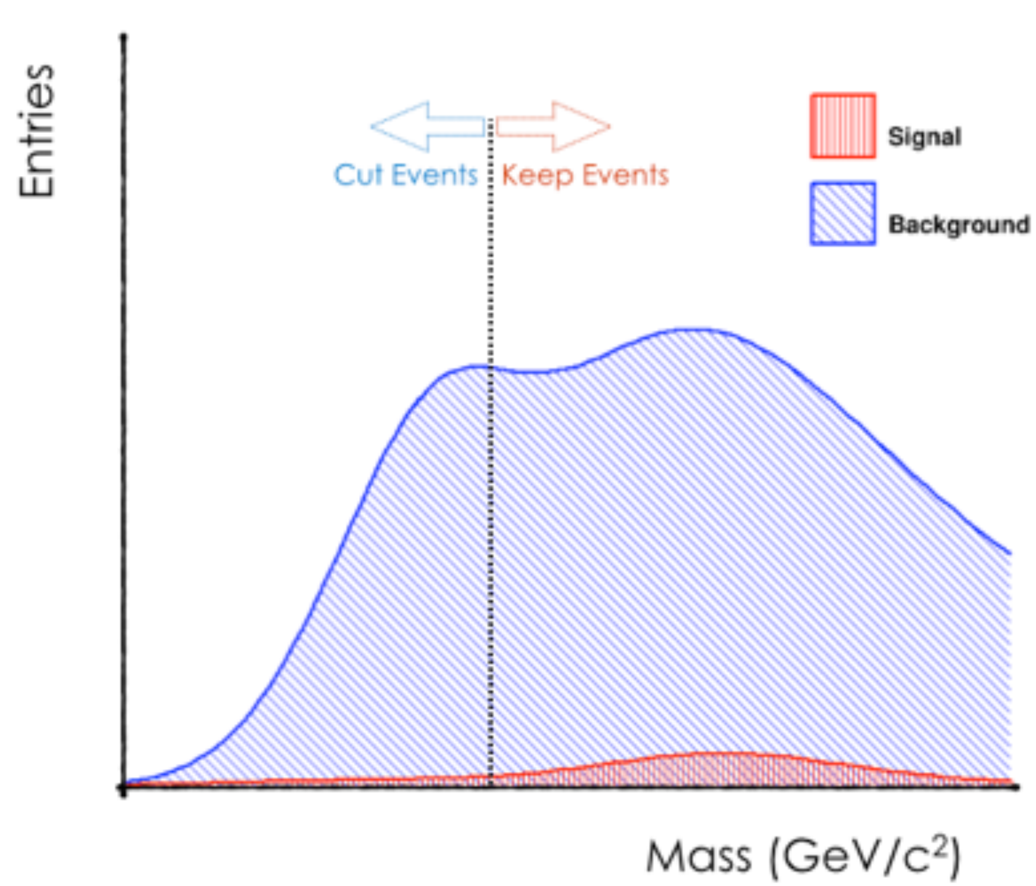
~ 2100 physicists
37 countries
167 universities/labs

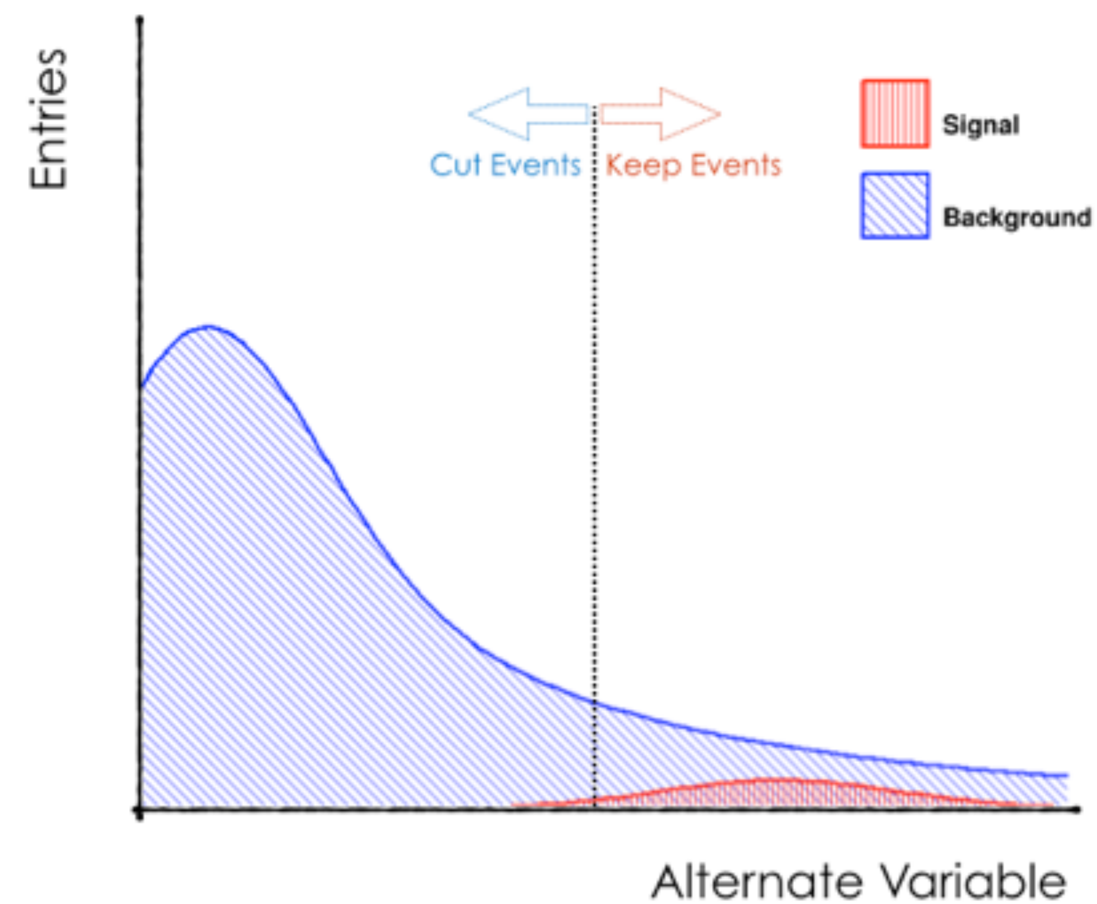




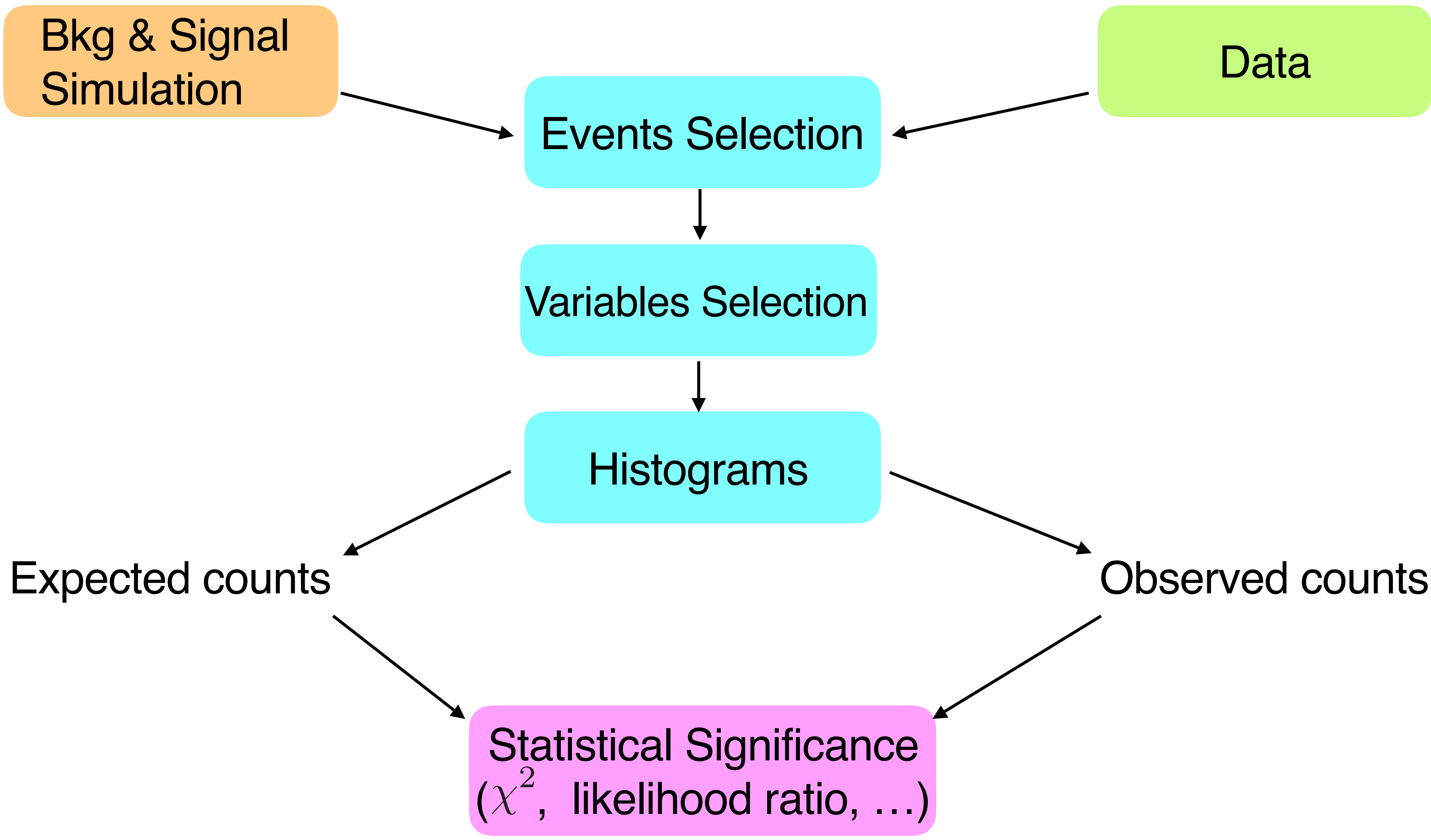


> Signal vs Background

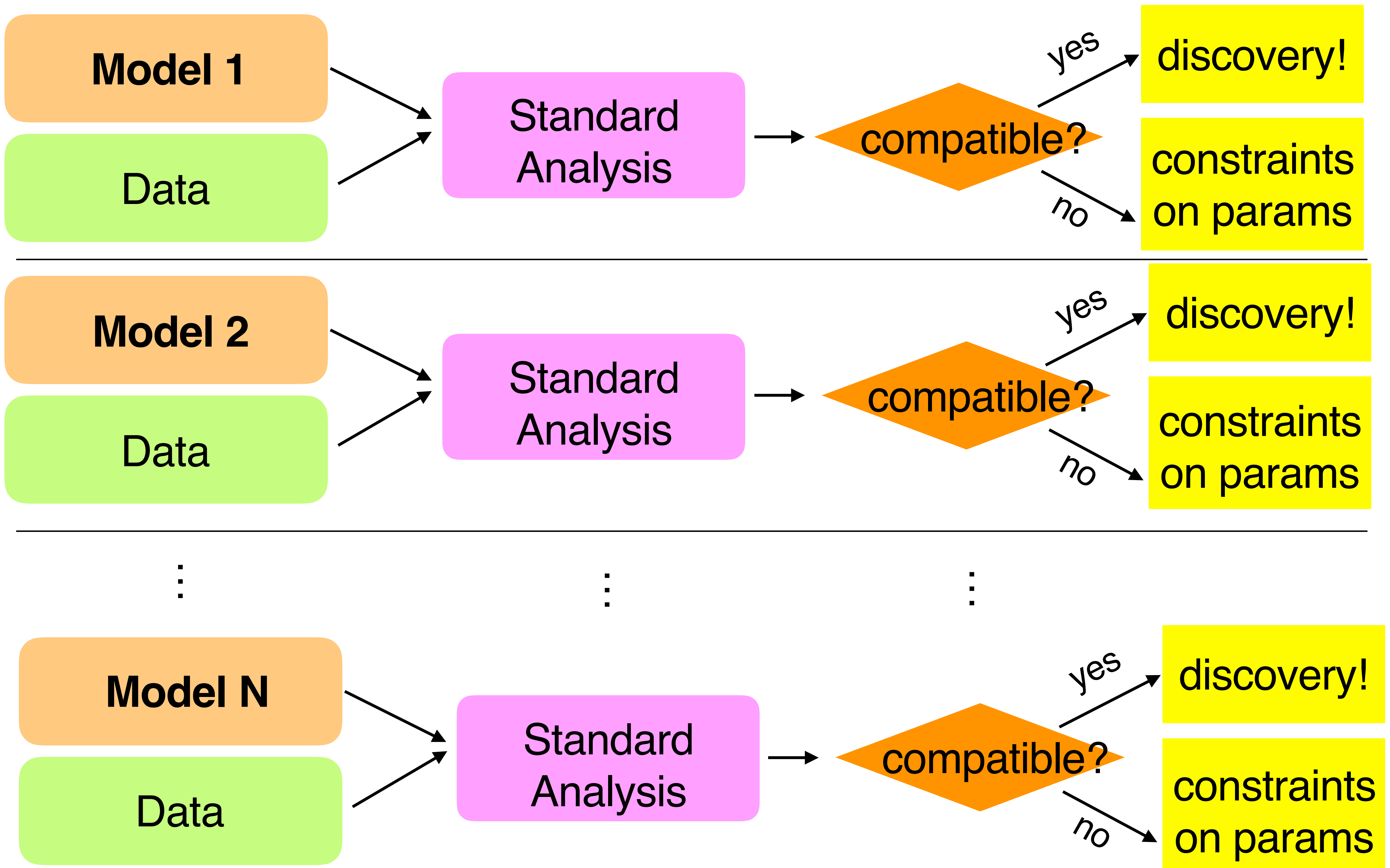




> Standard Analysis Pipeline



> Standard Analysis Pipeline



> New Physics ?

Searches for New Physics Beyond the Standard Model
have been negative so far…

MAYBE:

1. New Physics is not accessible by LHC

new particles are too light/heavy
or interacting too weakly

2. We have not explored all the possibilities

new physics may be buried under large bkg
or hiding behind unusual signatures

> New Physics ?

“Don’t want to miss a thing” (in data)

closer look at currently available data
get ready for upcoming data from next Run of LHC

Model-independent search

searches for specific models may be insensitive
to unexpected / unknown / anomalous processes

> Outline

1. Machine Learning in Science

2. Open problems in High-Energy Physics (HEP)

3. Statistical test of dataset compatibility

4. Applications to HEP

> New Statistical Test

[DS, Jacques - 1807.06038]

Want a statistical test for NP which is:

1. **model-independent:**

   no assumption about underlying physical model to intepret data

   → more general

2. **non-parametric:**

   compare two samples as a whole (not just their means, etc.)

   → fewer assumptions, no max likelihood estim.

3. **un-binned:**

   high-dim feature space partitioned without rectangular bins

   → retain full multi-dim info of data

[a.k.a. "homogeneity test"]

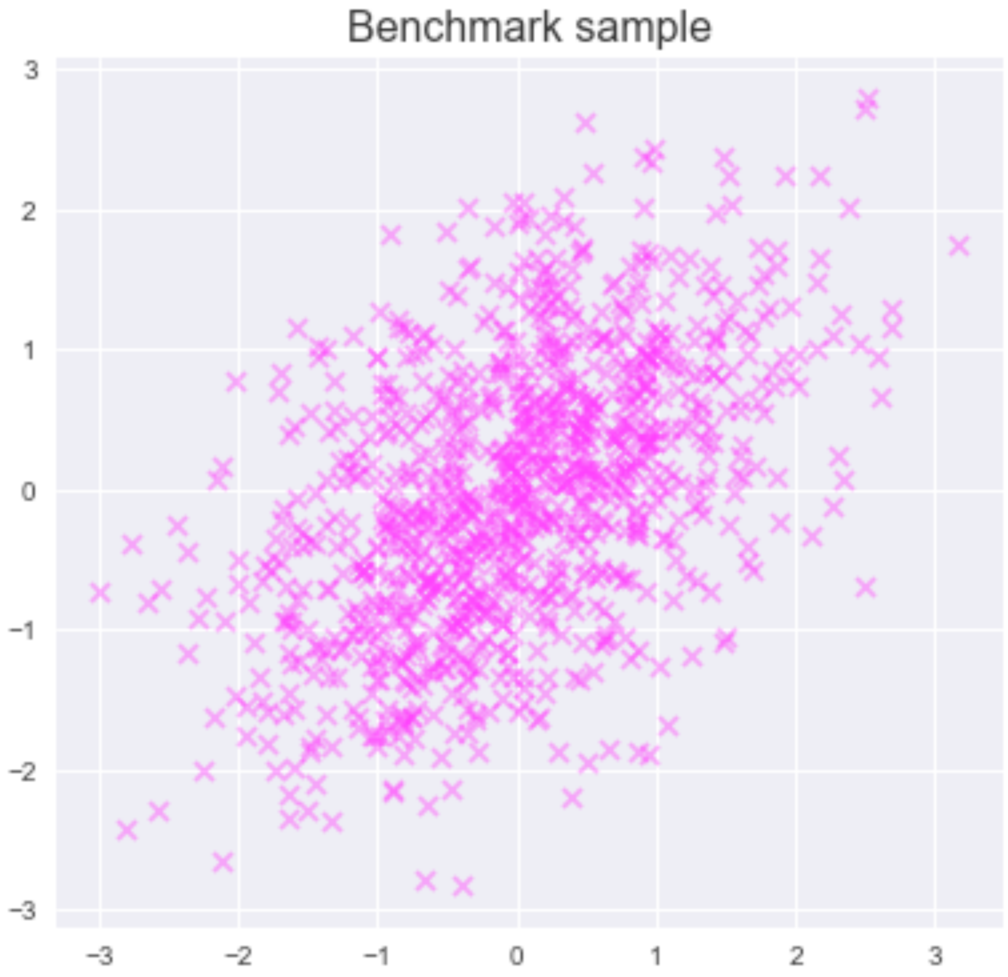
Two sets:

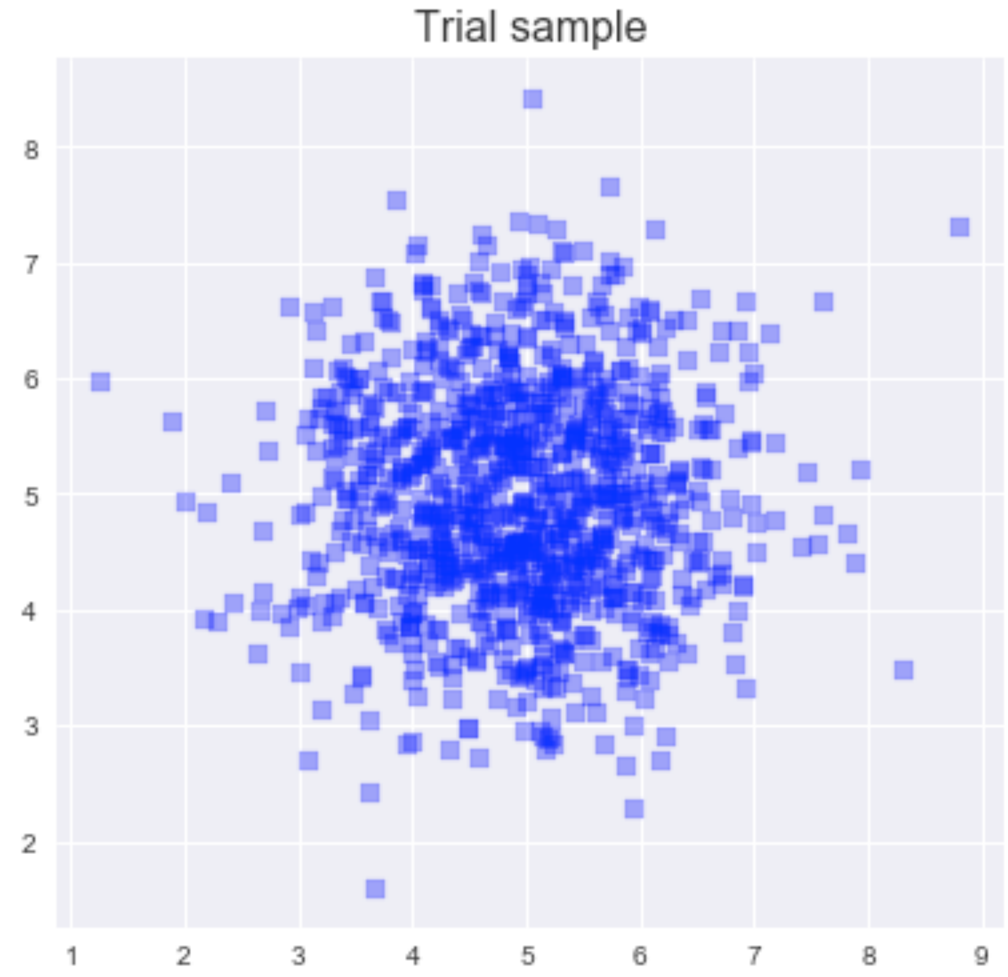
Trial: $\mathcal{T} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_T}\} \overset{\text{iid}}{\sim} p_T$

Benchmark: $\mathcal{B} = \{\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_{N_B}\} \overset{\text{iid}}{\sim} p_B$

$\boldsymbol{x}_i, \boldsymbol{x}'_i \in \mathbb{R}^D$

probability distributions p_B, p_T unknown





e.g.: simulated SM bkg

real measured data

Two sets:

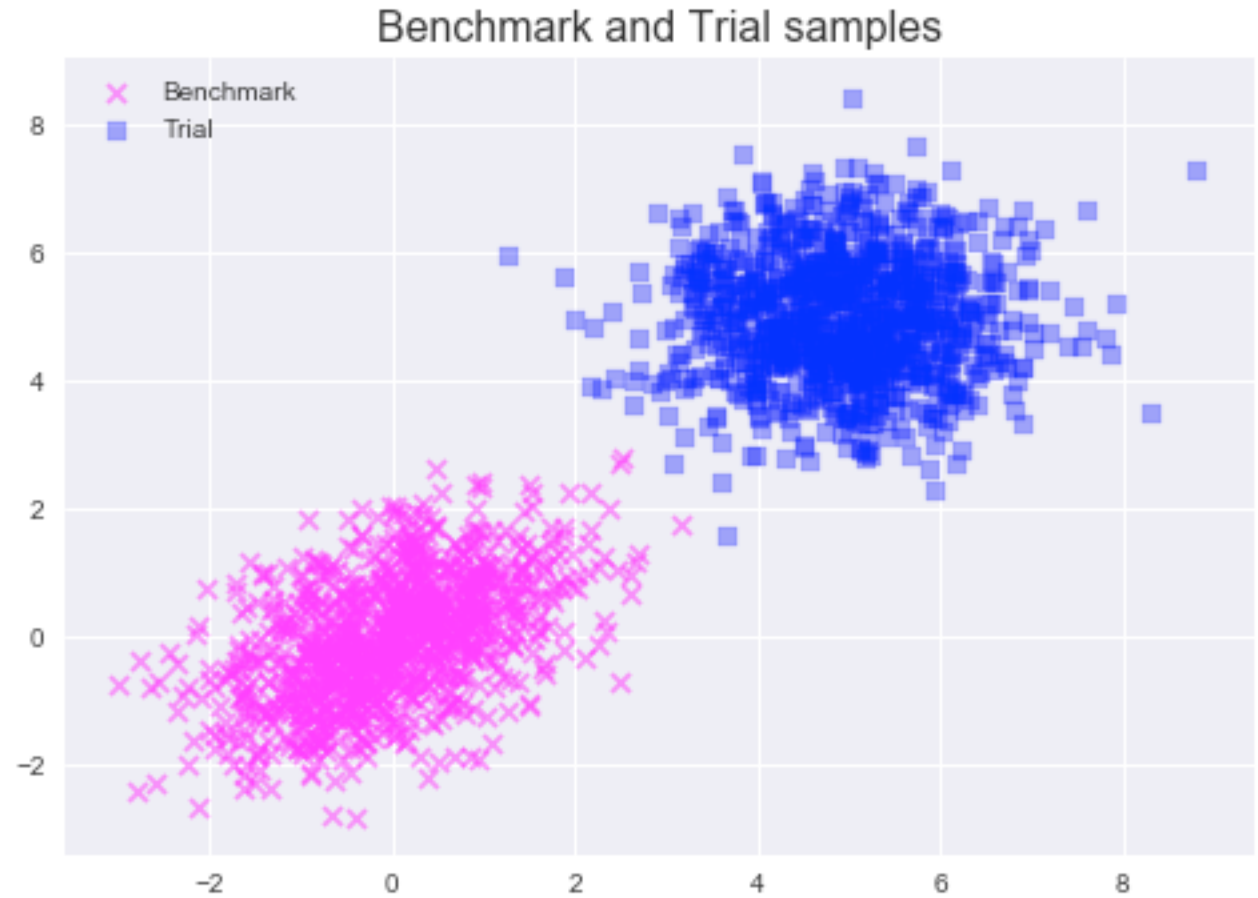
Trial: $\mathcal{T} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_T}\} \overset{\text{iid}}{\sim} p_T$

Benchmark: $\mathcal{B} = \{\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_{N_B}\} \overset{\text{iid}}{\sim} p_B$

$\boldsymbol{x}_i, \boldsymbol{x}'_i \in \mathbb{R}^D$

probability distributions p_B, p_T unknown

Are *B,T* drawn from the same prob. distribution?



easy…

> Two-sample Test

Two sets:

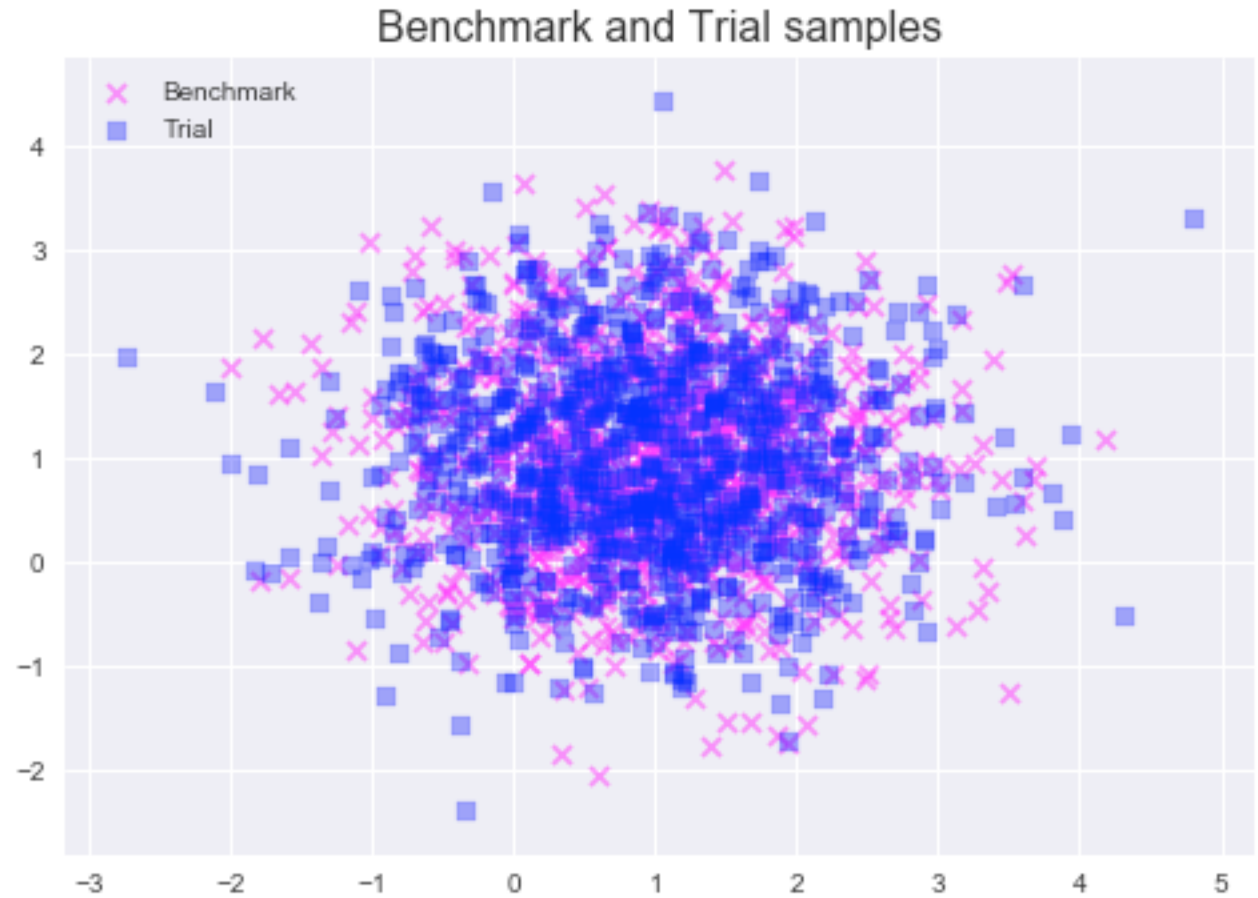
Trial: $\mathcal{T} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_T}\} \overset{\text{iid}}{\sim} p_T$

Benchmark: $\mathcal{B} = \{\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_{N_B}\} \overset{\text{iid}}{\sim} p_B$

$\boldsymbol{x}_i, \boldsymbol{x}'_i \in \mathbb{R}^D$

probability distributions p_B, p_T unknown

Are *B,T* drawn from the same prob. distribution?



… hard!

> Two-sample Test

Why is it important?

- decide whether two datasets can be analyzed jointly
- find anomalous data points (outliers)
- detect changes in the underlying distributions
- detect events in streams of data (time-series data)
- **check if data are compatible with expectations**
- …

> Two-sample Test

RECIPE:

1. Density Estimator

→ reconstruct PDFs from samples

2. Test Statistic (TS)

→ measure "distance" between PDFs

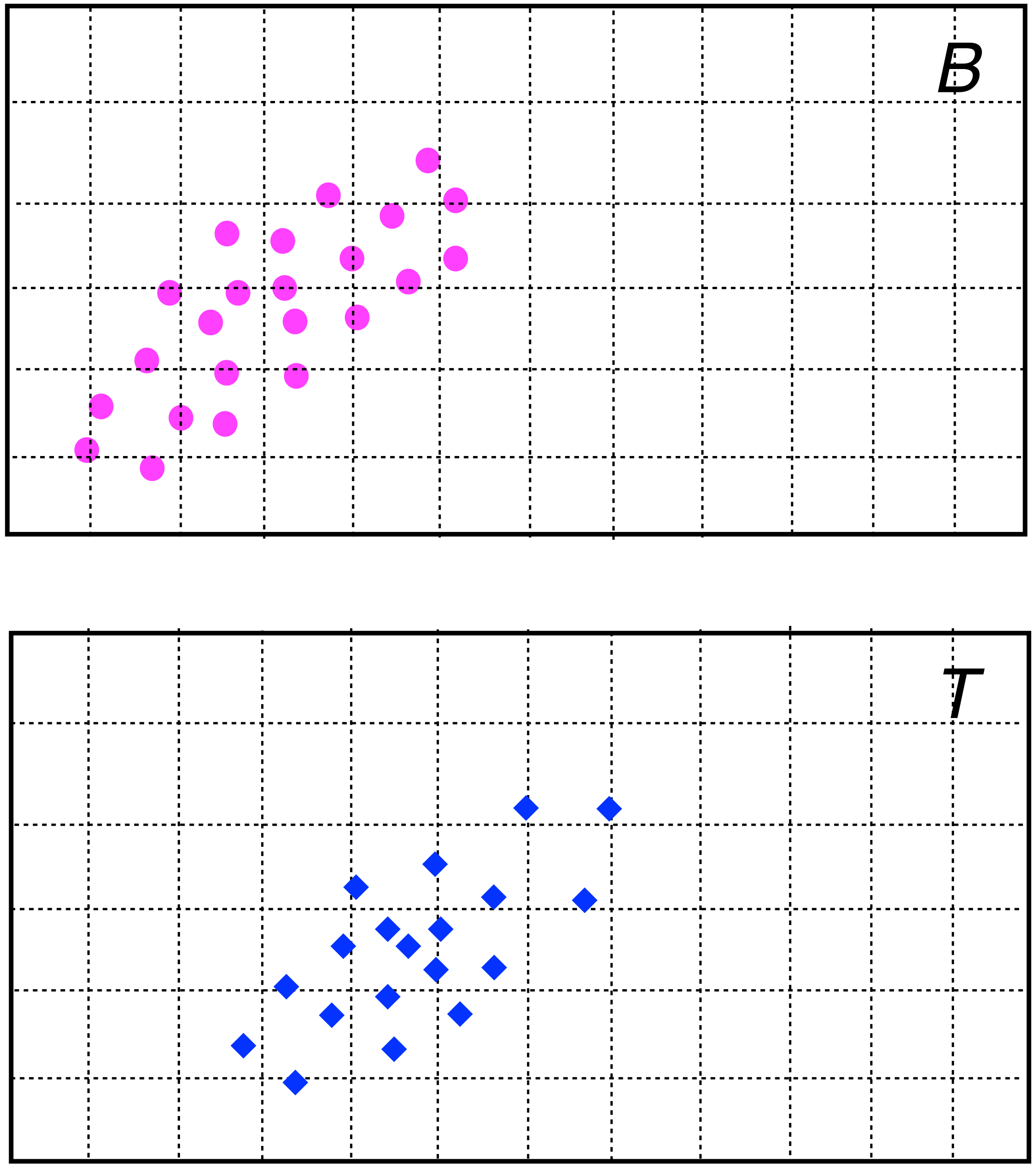
3. TS distribution

→ associate probabilities to TS under null hypothesis H_0: $p_B = p_T$

4. *p* -value

→ accept/reject H_0

> 1. Density Estimator



Divide the space in squared bins?

- ✓ easy
- ✓ can use simple statistics (e.g. χ^2)
- ✗ hard/slow/impossible in high-*D*

Need un-binned multivariate approach

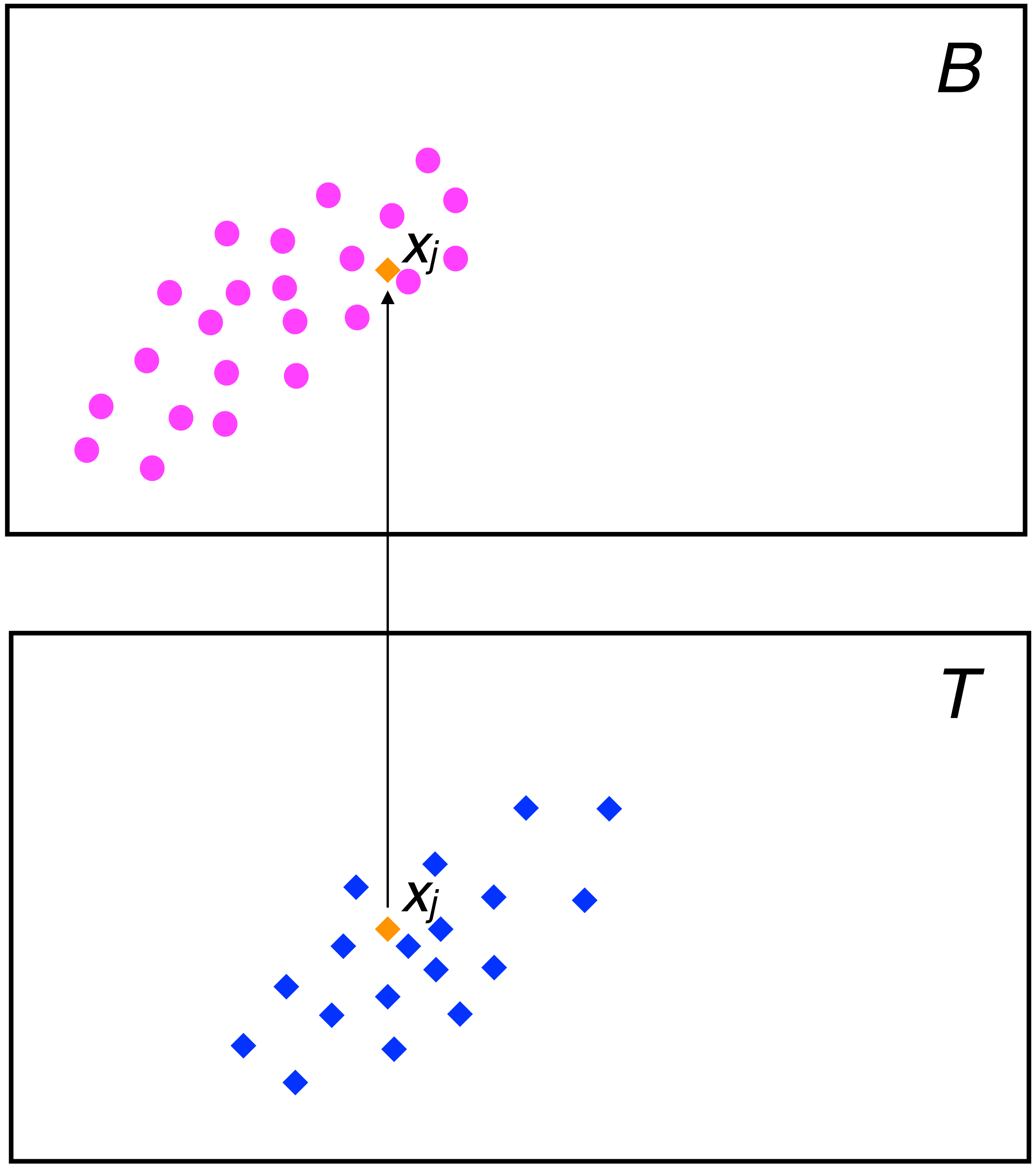
Find PDFs *estimators*: $\hat{p}_B(\boldsymbol{x}), \hat{p}_T(\boldsymbol{x})$
e.g. based on densities of points:

$$\hat{p}_{B,T}(\boldsymbol{x}) = \frac{\rho_{B,T}(\boldsymbol{x})}{N_{B,T}}$$

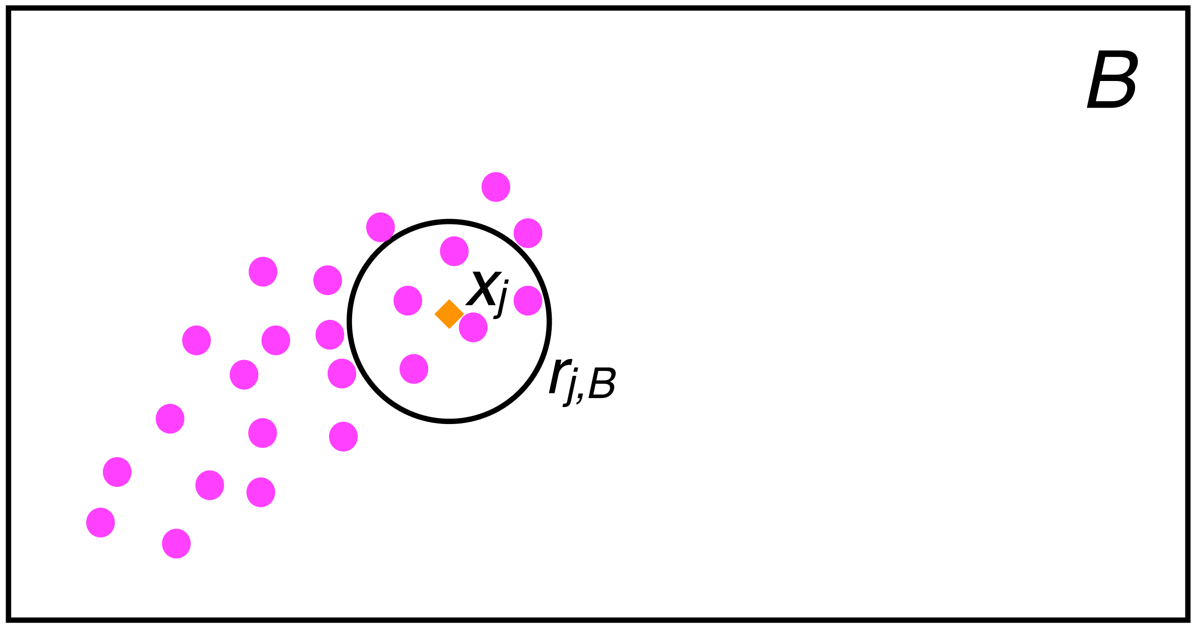
Nearest Neighbors!

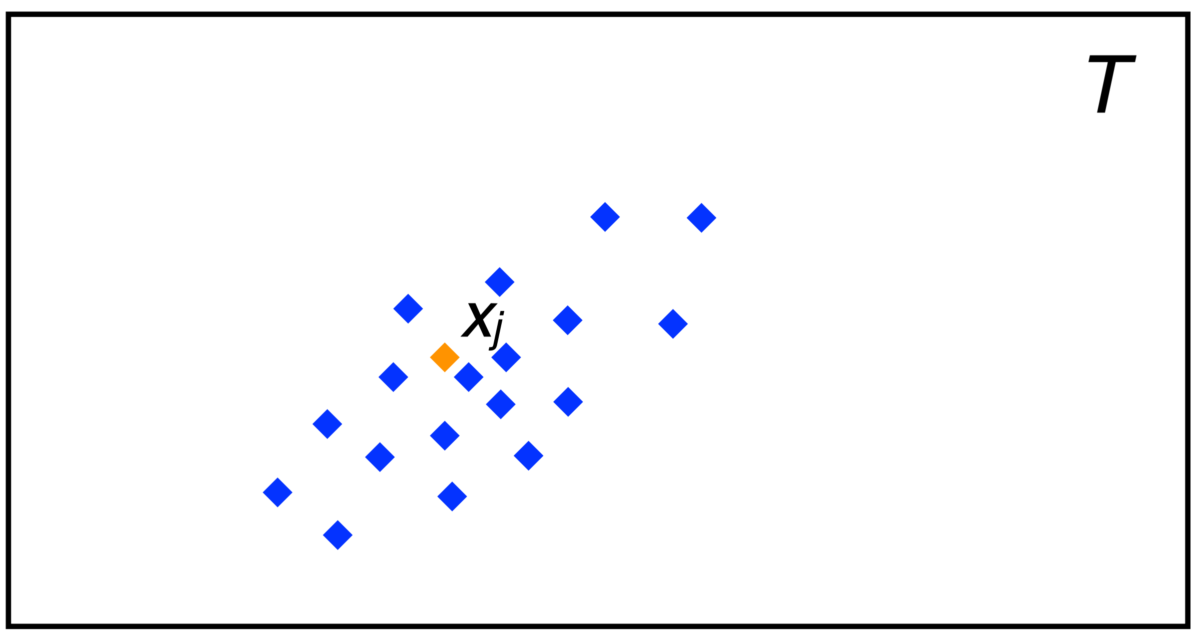
[Schilling - 1986][Henze - 1988]
[Wang et al. - 2005,2006]
[Dasu et al. - 2006][Perez-Cruz - 2008]
[Sugiyama et al. - 2011][Kremer et al, 2015]

> 1. Density Estimator



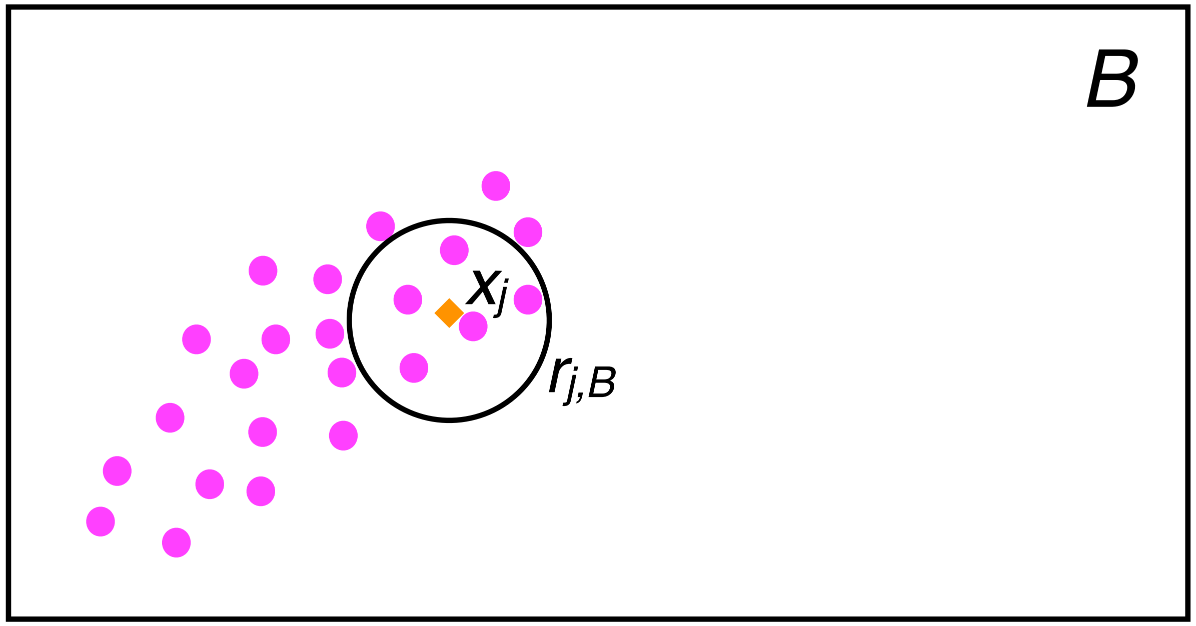
- Fix integer K.
- Choose query point x_j in *T* and draw it in *B.*

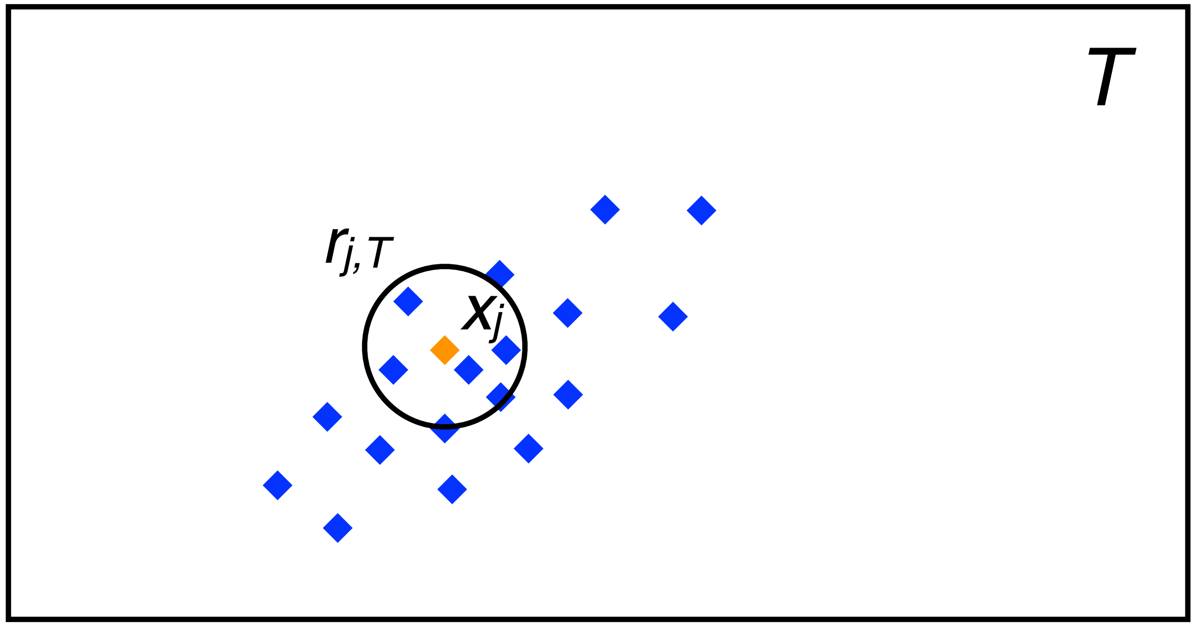




- Fix integer K.
- Choose query point x_j in T and draw it in B.
- Find the distance $r_{j,B}$ of the Kth-NN of x_j in B.

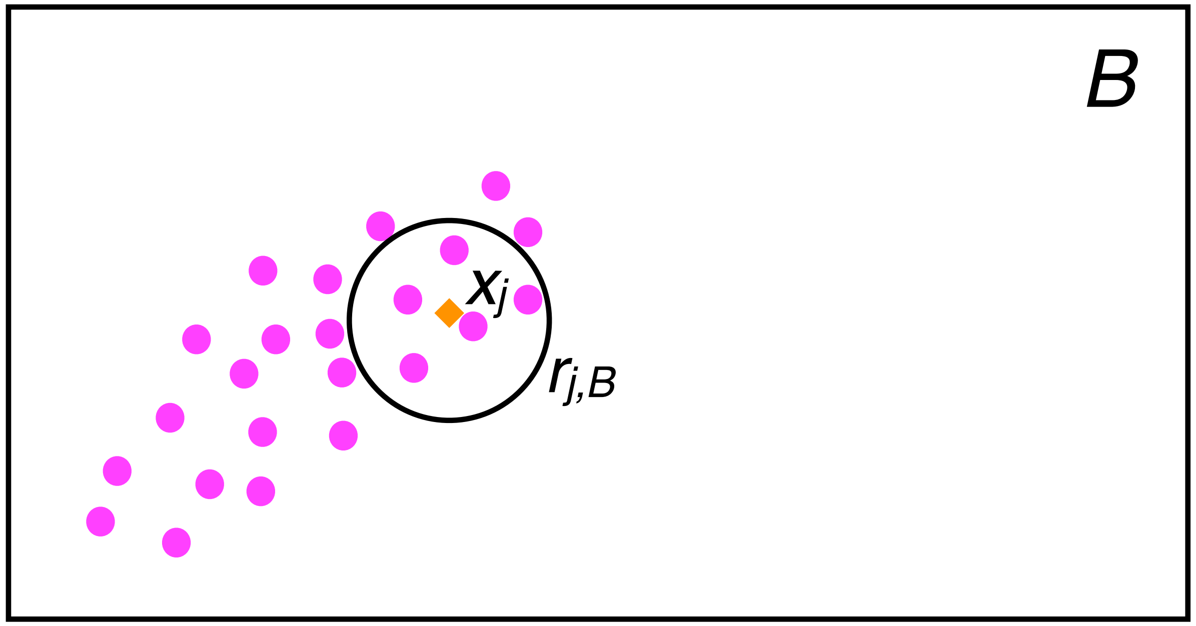
> 1. Density Estimator

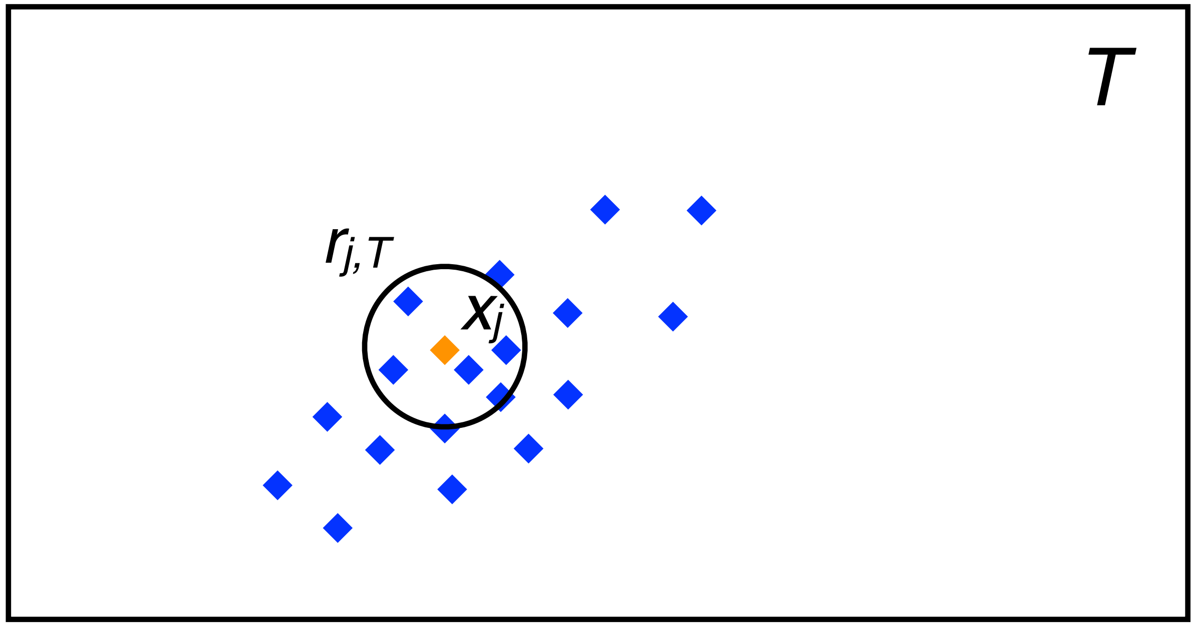




- Fix integer K.
- Choose query point x_j in T and draw it in B.
- Find the distance $r_{j,B}$ of the Kth-NN of x_j in B.
- Find the distance $r_{j,T}$ of the Kth-NN of x_j in T.

> 1. Density Estimator





- Fix integer K.
- Choose query point x_j in *T* and draw it in *B.*
- Find the distance $r_{j,B}$ of the Kth-NN of x_j in *B.*
- Find the distance $r_{j,T}$ of the Kth-NN of x_j in *T.*
- Estimate PDFs:

$$\hat{p}_B(\boldsymbol{x}_j) = \frac{K}{N_B} \frac{1}{\omega_D r_{j,B}^D}$$

$$\hat{p}_T(\boldsymbol{x}_j) = \frac{K}{N_T - 1} \frac{1}{\omega_D r_{j,T}^D}$$

> 2. Test Statistic

- Measure of the “distance” between 2 PDFs

- Define **Test Statistic**:
  (detect under-/over-densities)

$$\boxed{\mathrm{TS}(\mathcal{T}) \equiv \frac{1}{N_T} \sum_{j=1}^{N_T} \log \frac{\hat{p}_T(\boldsymbol{x}_j)}{\hat{p}_B(\boldsymbol{x}_j)}}$$

- Related to Kullback-Leibler divergence as: $\mathrm{TS}(\mathcal{T}) = \hat{D}_{\mathrm{KL}}(\hat{p}_T || \hat{p}_B)$

$$D_{\mathrm{KL}}(p||q) \equiv \int_{\mathbb{R}^D} p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} d\boldsymbol{x}$$

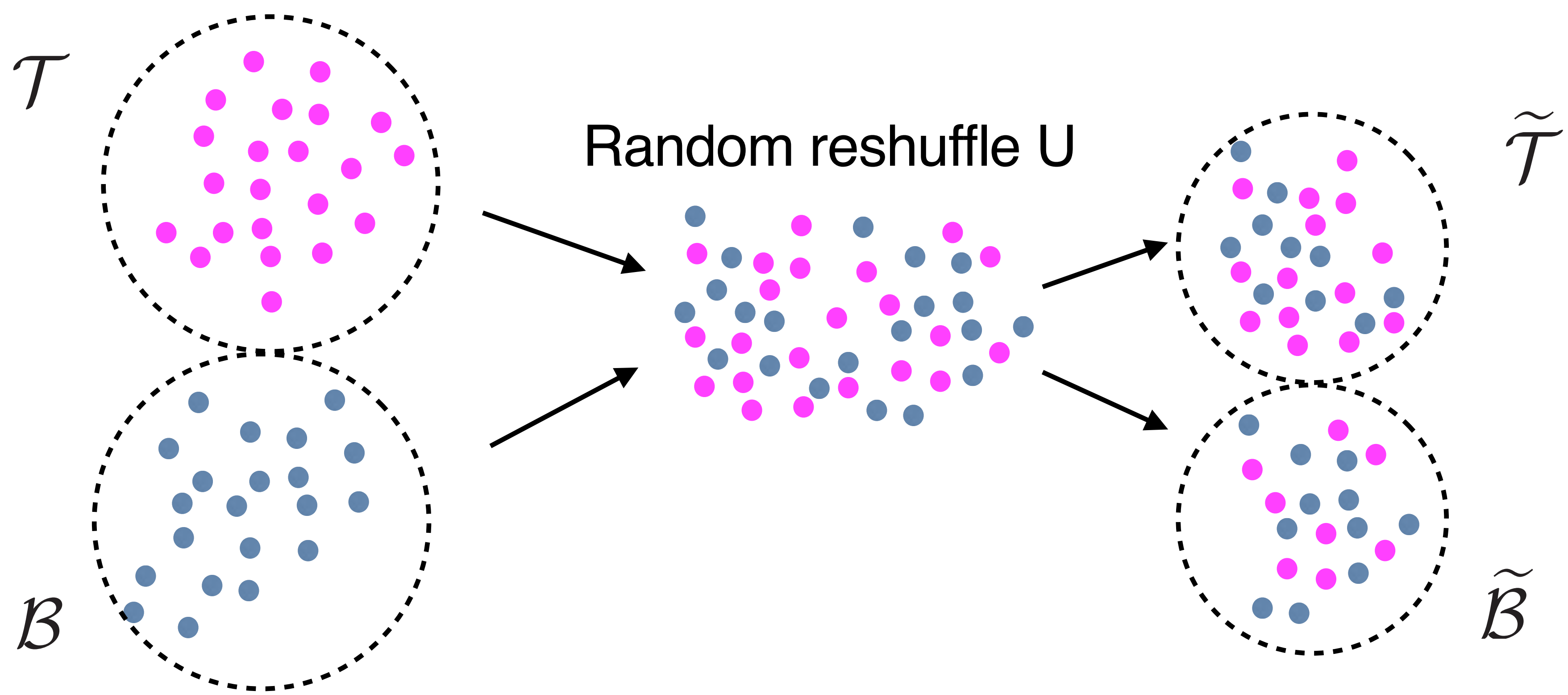
- From NN-estimated PDFs: $\mathrm{TS}(\mathcal{T}) = \frac{D}{N_T} \sum_{j=1}^{N_T} \log \frac{r_{j,B}}{r_{j,T}} + \log \frac{N_B}{N_T - 1}$

- **Theorem:** this estimator converges to $D_{\mathrm{KL}}(p_B || p_T)$,
  in large sample limit

[Wang et al. - 2005,2006]

How is TS distributed? **Permutation test!**

Assume $p_B = p_T$. Union set: $\mathcal{U} = \mathcal{T} \cup \mathcal{B}$

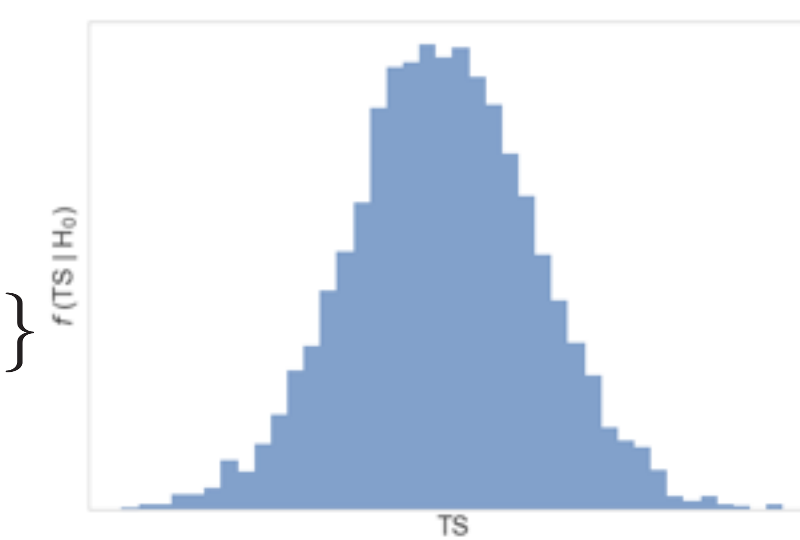


Compute the test statistic TS_n on: $(\tilde{\mathcal{B}}, \tilde{\mathcal{T}})$

Repeat many times.

Distribution of TS under H_0: $f(\mathrm{TS}|H_0) \leftarrow \{\mathrm{TS}_n\}$

[asymptotically normal with zero mean]



> 4. *p*-value

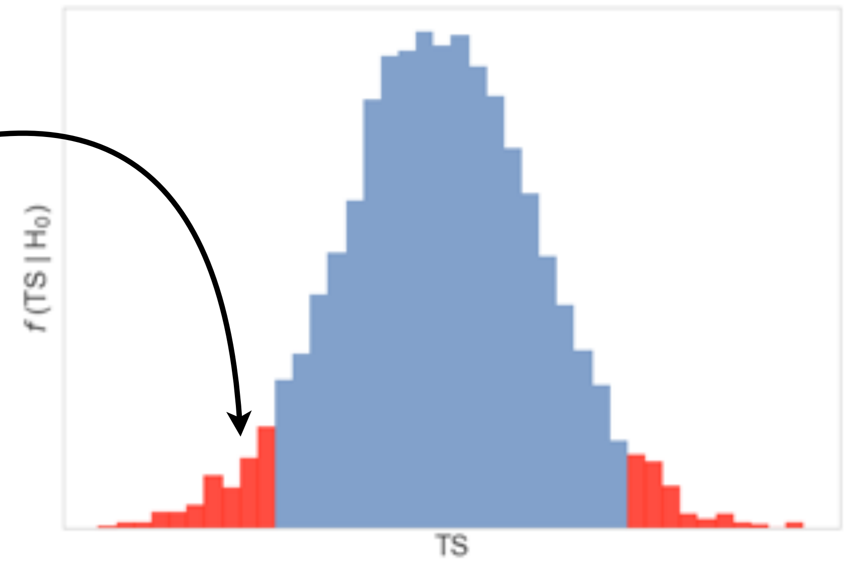
- $\hat{\mu}, \hat{\sigma}$: mean,variance of TS distribution $f(\mathrm{TS}|H_0)$

- Standardize the TS: $\mathrm{TS} \to \mathrm{TS}' \equiv \dfrac{\mathrm{TS} - \hat{\mu}}{\hat{\sigma}}$

- TS' distributed according to $f'(\mathrm{TS}'|H_0) = \hat{\sigma} f(\hat{\mu} + \hat{\sigma}\mathrm{TS}'|H_0)$

- Two-sided *p*-value:

$$p = 2\int_{|\mathrm{TS}'_{\mathrm{obs}}|}^{+\infty} f'(\mathrm{TS}'|H_0) d\mathrm{TS}'$$

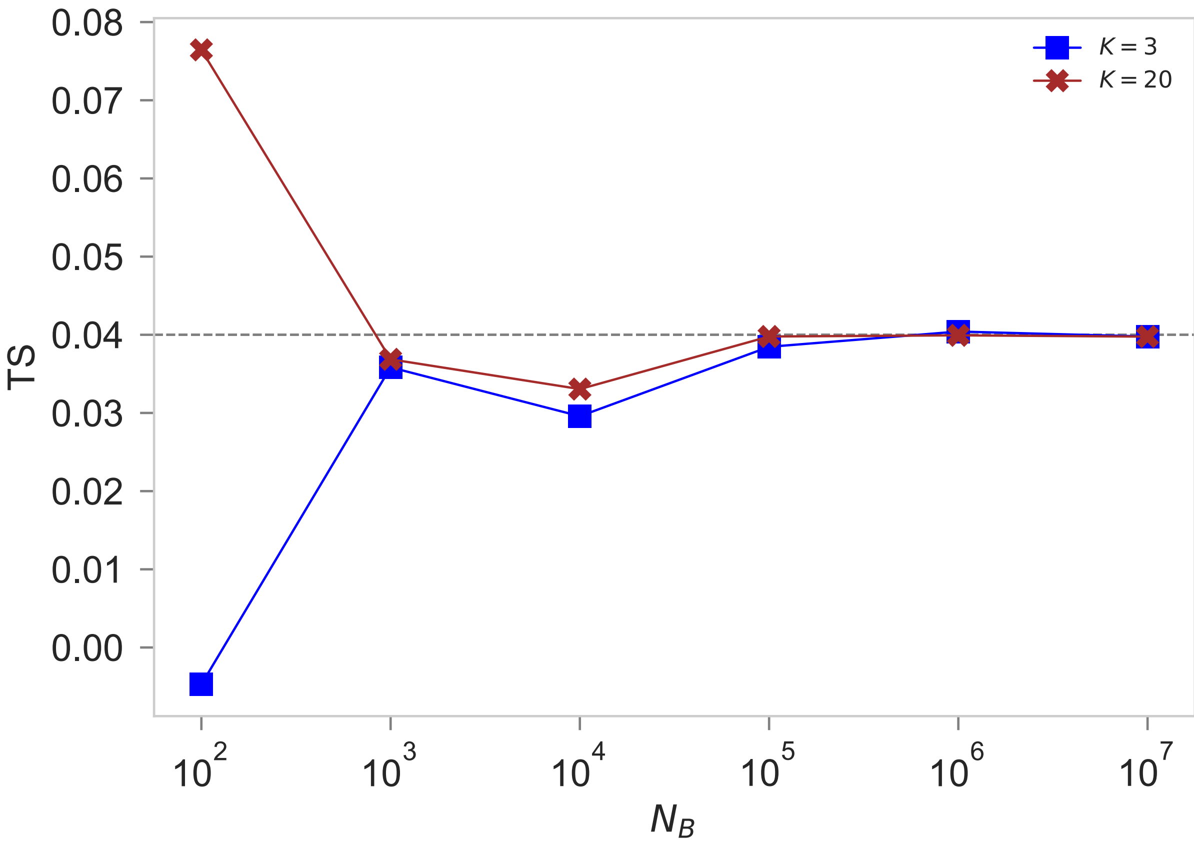


- Equivalent significance: $Z \equiv \Phi^{-1}(1 - p/2)$

> Gaussian Example

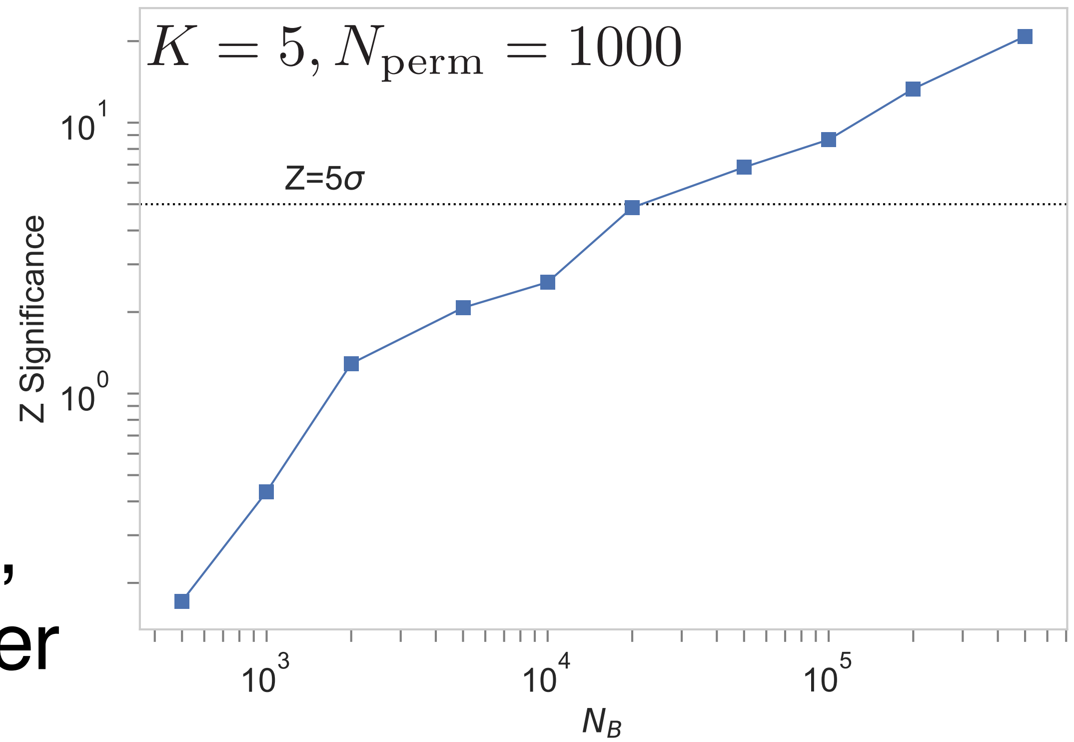
$$p_B = \mathcal{N}(\boldsymbol{\mu}_B, \Sigma_B) \qquad p_T = \mathcal{N}(\boldsymbol{\mu}_T, \Sigma_T) \qquad \Sigma_B = \Sigma_T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{\mu}_B = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix} \qquad \boldsymbol{\mu}_T = \begin{pmatrix} 1.2 \\ 1.2 \end{pmatrix}$$



exact KL divergence

$$\boldsymbol{\mu}_B = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix} \qquad \boldsymbol{\mu}_T = \begin{pmatrix} 1.15 \\ 1.15 \end{pmatrix}$$



more data, more power

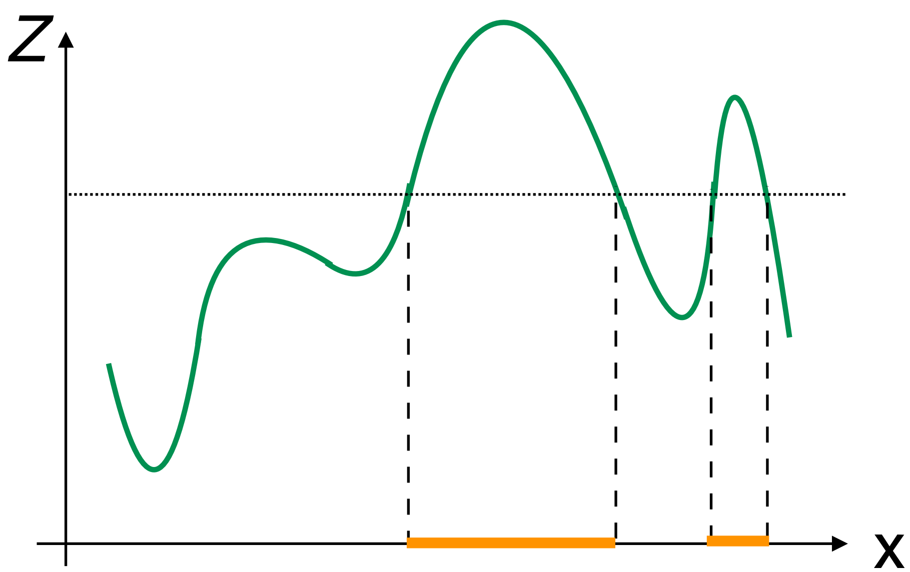
> Where are the discrepancies?

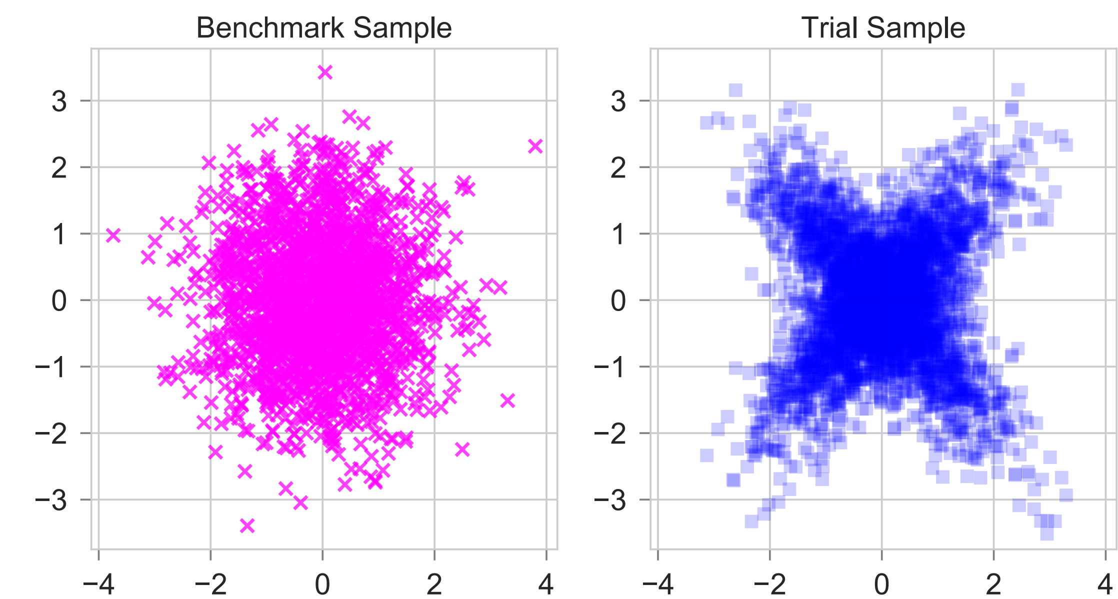
Bonus: Characterize regions with significant discrepancies

1. "Score" field over *T*: $Z(\boldsymbol{x}_j) \equiv \dfrac{u(\boldsymbol{x}_j) - \bar{u}}{\sigma_u}$

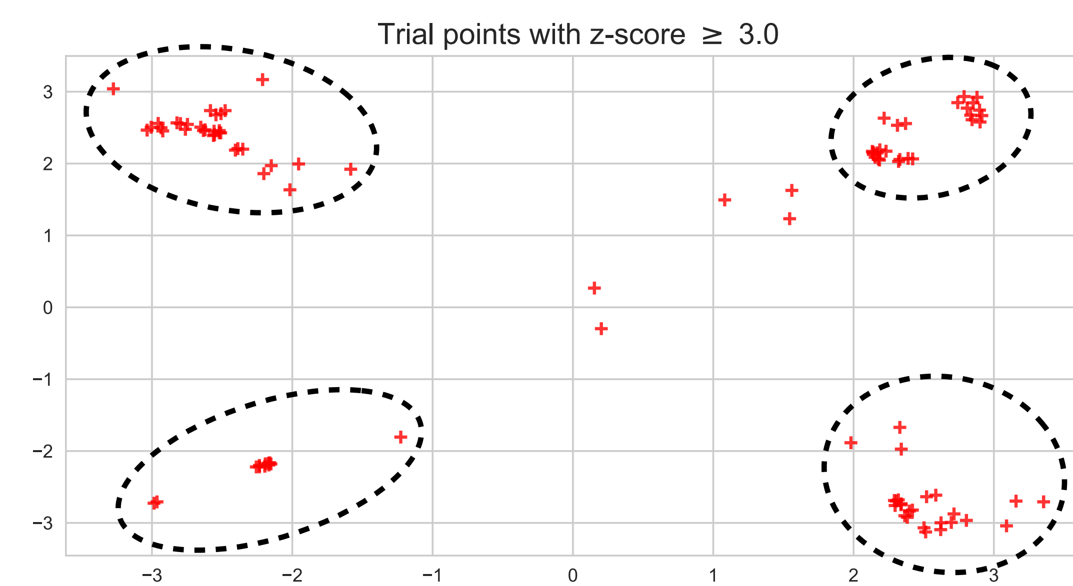
with: $u(\boldsymbol{x}_j) \equiv \log \dfrac{r_{j,B}}{r_{j,T}}$

$$\mathrm{TS}_{\mathrm{obs}} = D\,\bar{u} + \mathrm{const}$$





2. Identify points where $Z(\boldsymbol{x}) > c$
They contribute the most to large $\mathrm{TS}_{\mathrm{obs}}$
→ high-discrepancy (anomalous) regions



3. Apply a clustering algorithm to group them

INPUT:

Trial sample: $\mathcal{T} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_T}\} \overset{\text{iid}}{\sim} p_T$

Benchmark sample: $\mathcal{B} = \{\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_{N_B}\} \overset{\text{iid}}{\sim} p_B$

K: number of nearest neighbors

N_{perm}: number of permutations

$\boldsymbol{x}_i, \boldsymbol{x}'_i \in \mathbb{R}^D$

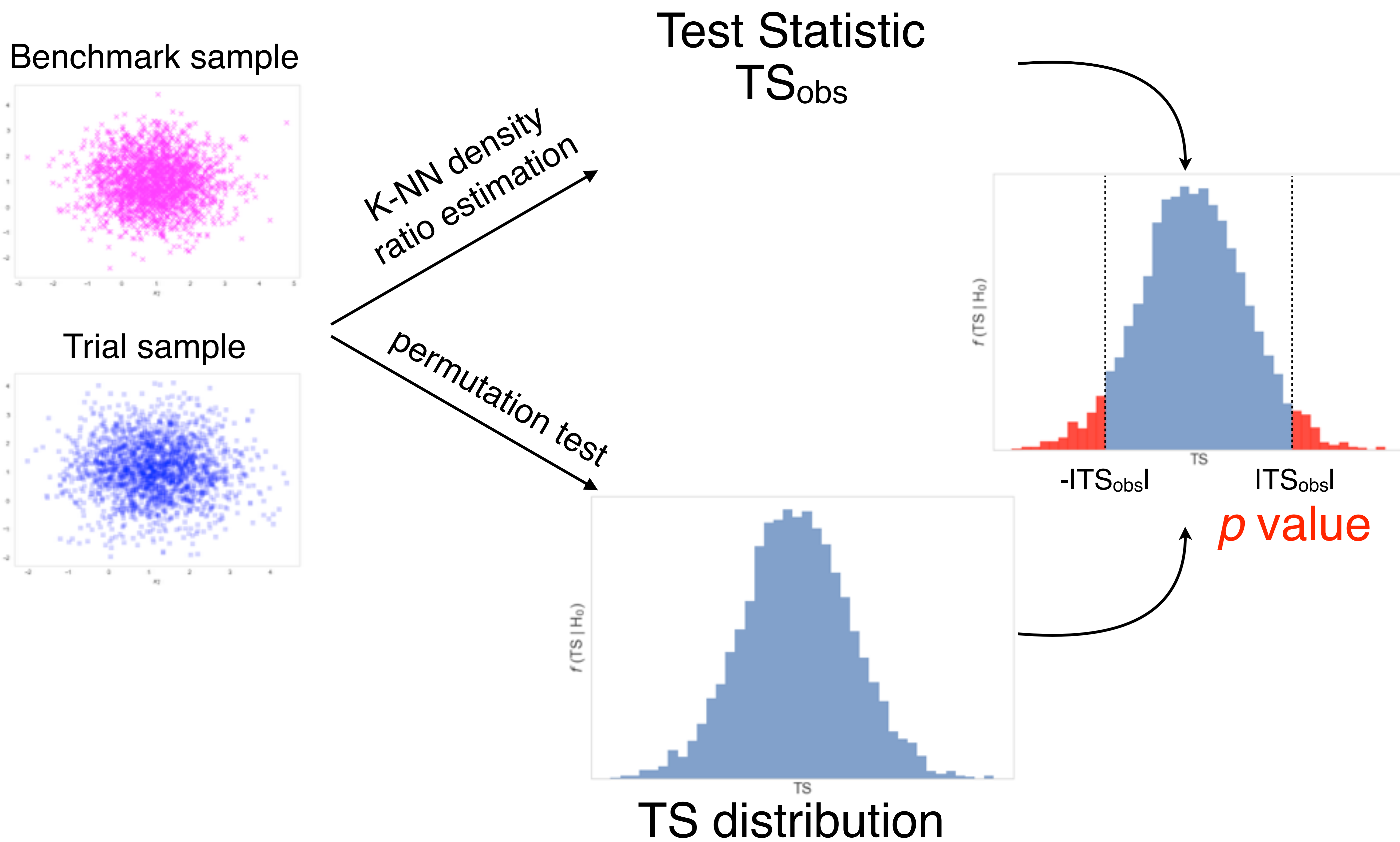
p_B, p_T unknown

OUTPUT:

***p*-value of the null hypothesis H_0: $p_B = p_T$**

[check compatibility between 2 samples]

> NN2ST: Summary



Python code:
github.com/de-simone/NN2ST

Paper:
[DS, Jacques - 1807.06038]

> NN2ST: Summary

✓ general, model-independent
✓ fast, no optimization
✓ sensitive to unspecified signals
✓ useful when no variable can separate sig/bkg
✓ helps finding signal regions, optimal cuts, …

✘ need to run for each sample pair
✘ permutation test is bottleneck
✘ limited by sample accuracies

> Outline

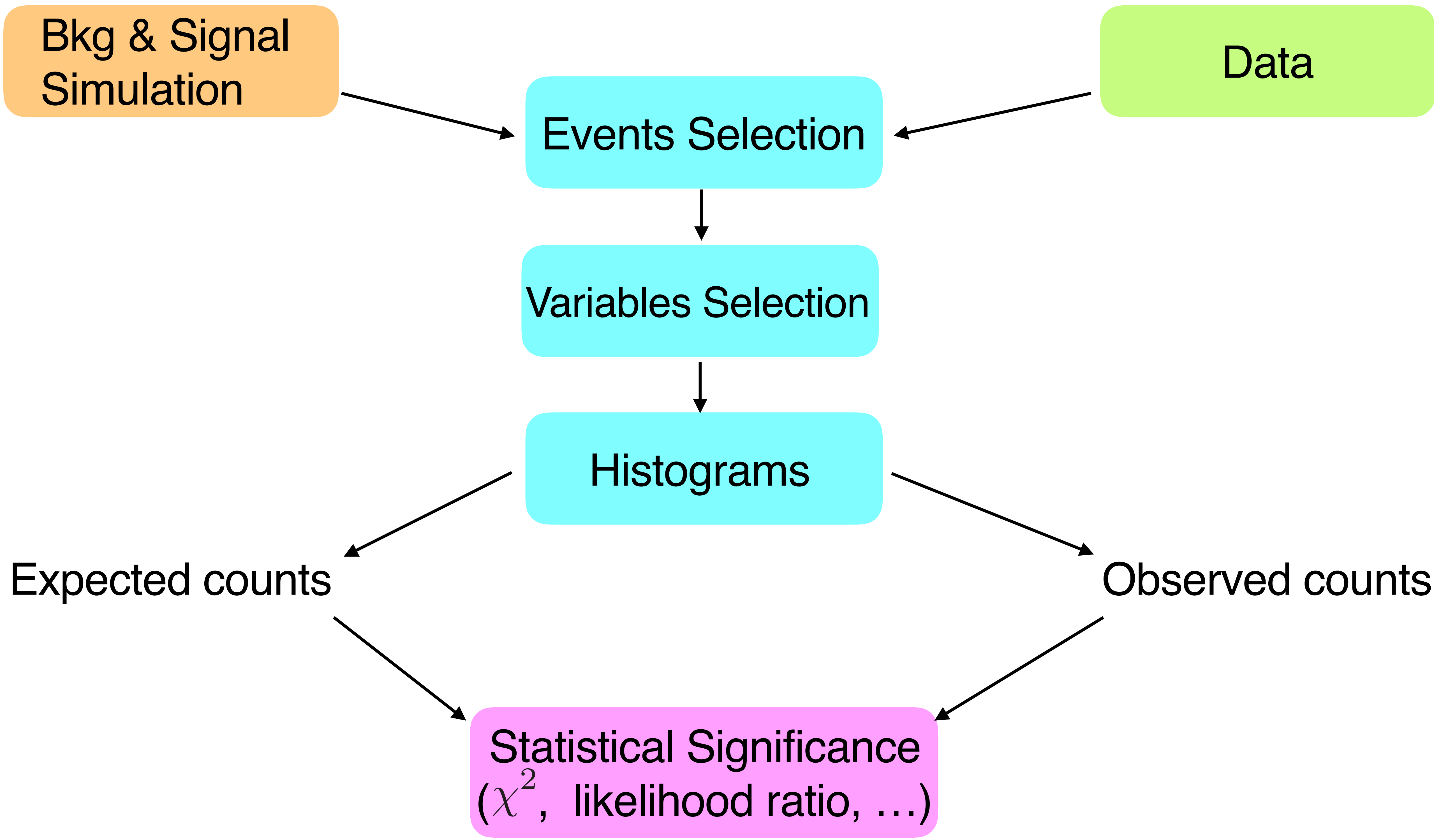
1. Machine Learning in Science

2. Open problems in High-Energy Physics (HEP)

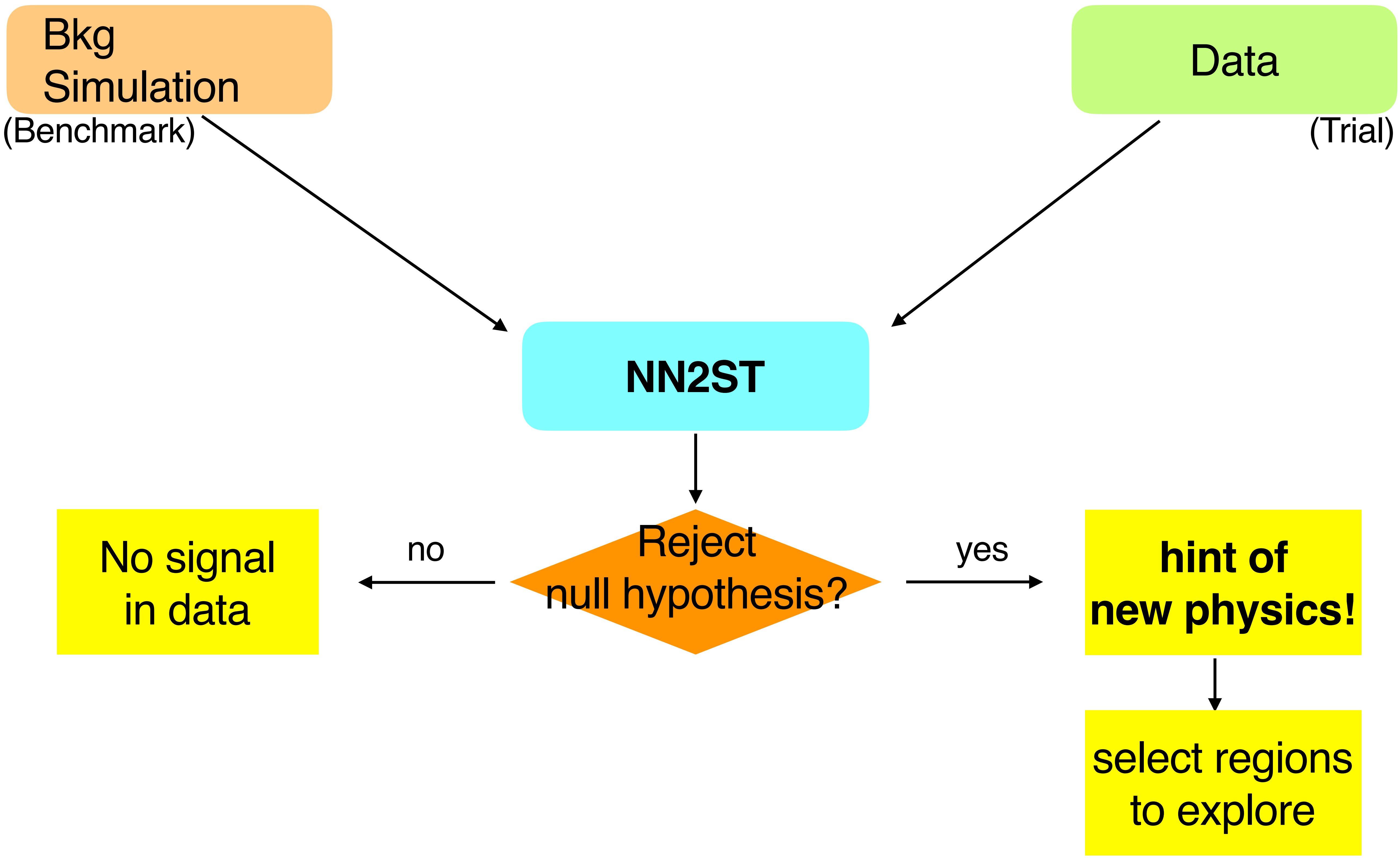
3. Statistical test of dataset compatibility

4. Applications to HEP

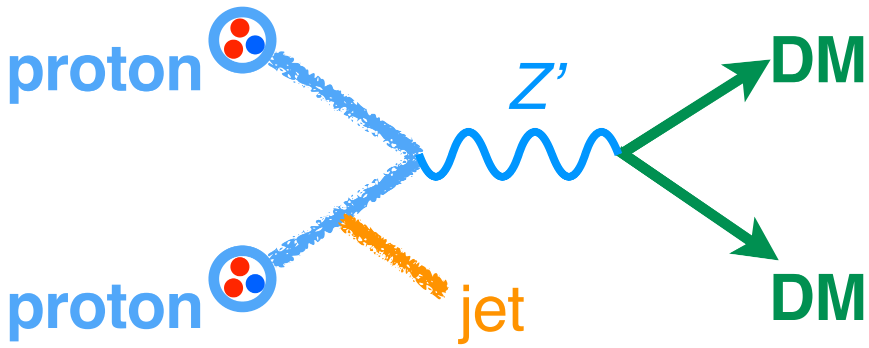
> Standard Analysis Pipeline



> Our Method

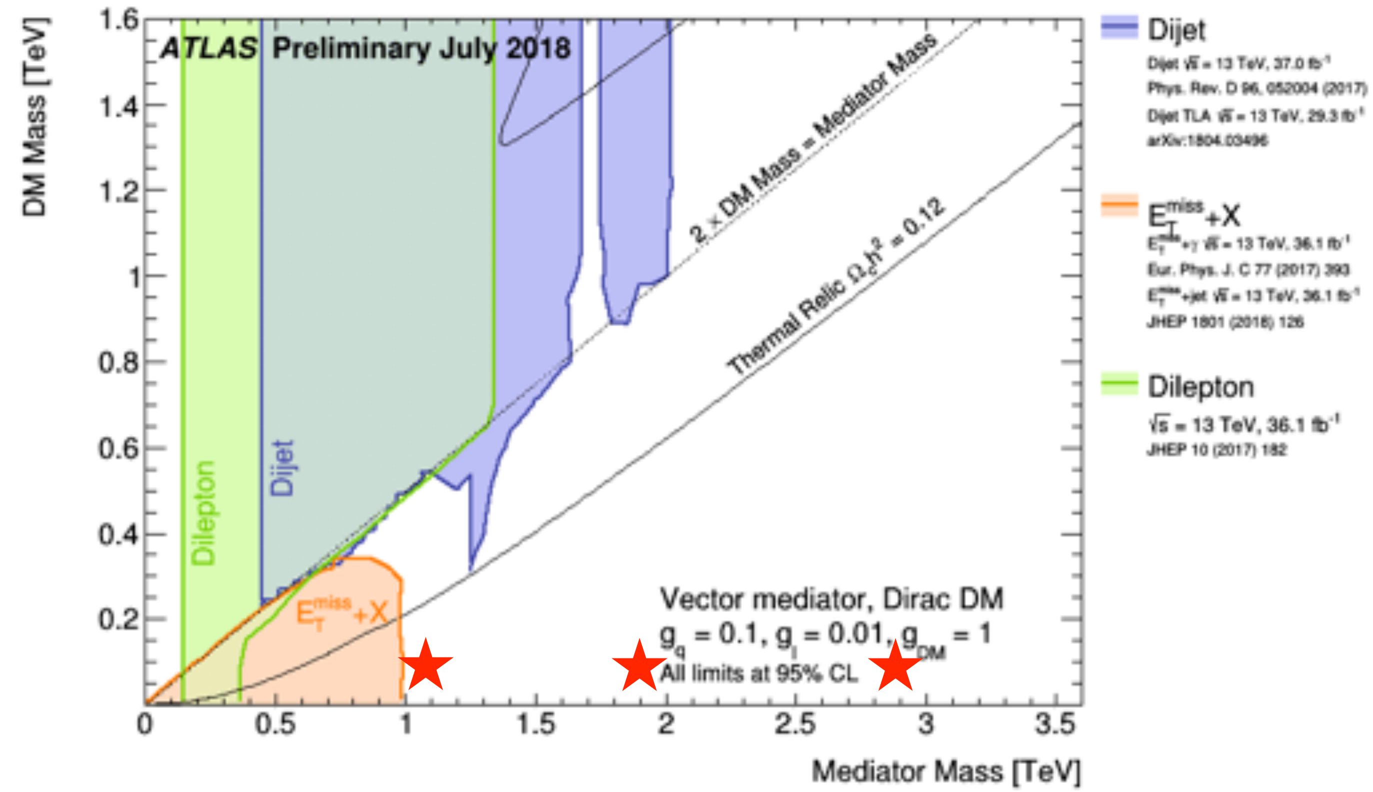

Bkg
Simulation
(Benchmark)
Data
(Trial)
NN2ST
Reject
null hypothesis?
no
No signal
in data
yes
hint of
new physics!
select regions
to explore

> DM search @ LHC



DM + Z'
vector mediator

m_{DM} = 100 GeV
$m_{Z'}$ = 1.2, 2, 3 TeV
g_{DM} = 1, g_q = 0.1
$\sqrt{s} = 13\ \mathrm{TeV}$



- "proof-of-principle" study
- bkg: $Z \to \nu\bar{\nu} + (1,2)\,j$
  sub-leading bkgs not included
- no full detector effects

Benchmark: **BKG_1**
Trials: **BKG_2 + SIG**
$K = 5$
$N_{perm} = 3000$

8 features:
- n. of jets
- p_T, η of 2 leading jets
- $E_T^{\rm miss}, H_T$
- $\Delta\phi_{E_T^{\rm miss}, j_1}$

B: **BKG$_1$** (20k events)
T1: **BKG$_2$** (20k events) **+ SIG$_1$** (2010 events)
T2: **BKG$_2$** (20k events) **+ SIG$_2$** (375 events)
T3: **BKG$_2$** (20k events) **+ SIG$_3$** (59 events)

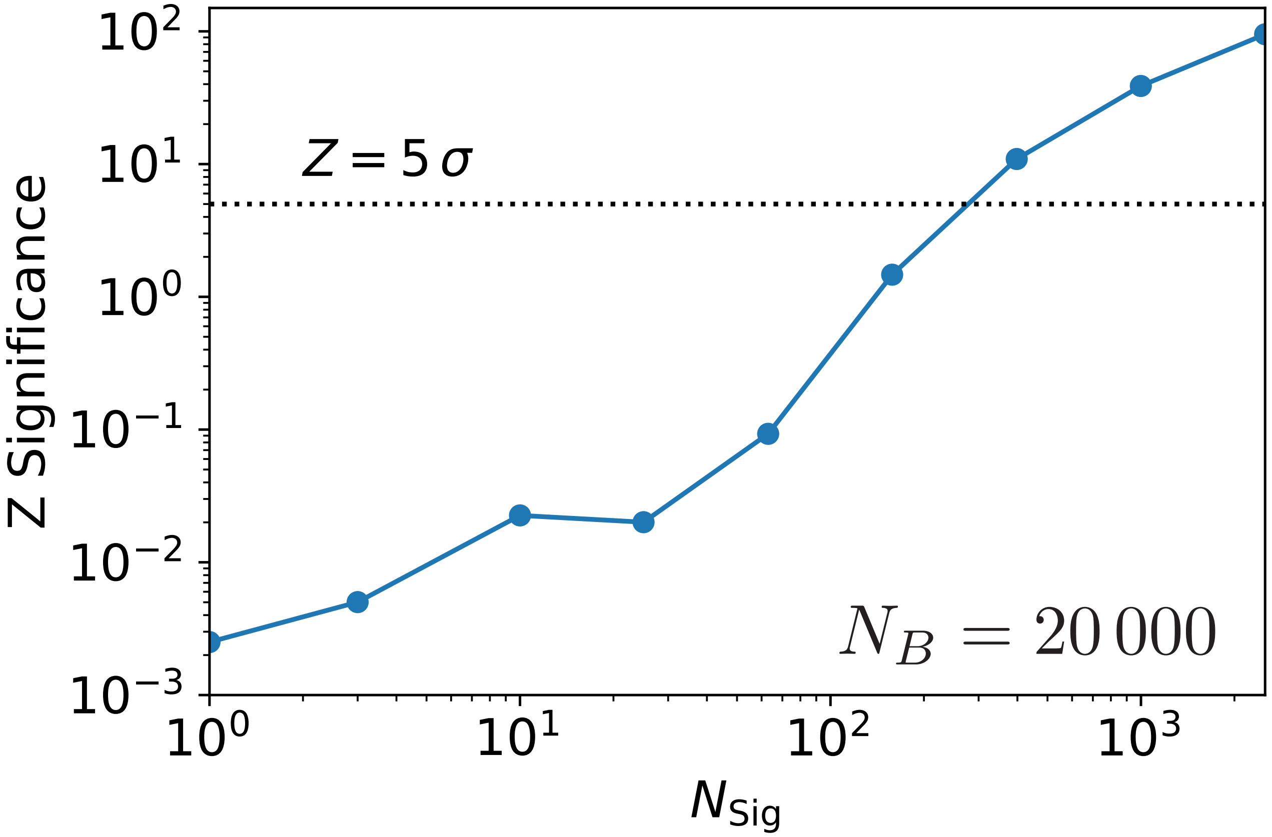
$$N_{\rm sig} = N_B \times \frac{\sigma_{\rm signal}}{\sigma_{\rm bkg}}$$

Sample	$M_{Z'}$	$\sigma_{\rm signal}$	Z
T1	1.2 TeV	20.4 pb	>15 σ
T2	2 TeV	3.8 pb	10 σ
T3	3 TeV	0.6 pb	0.13 σ

- in real world: expect degradation of results (uncertainties)
- **the method has value, it is worth exploring**

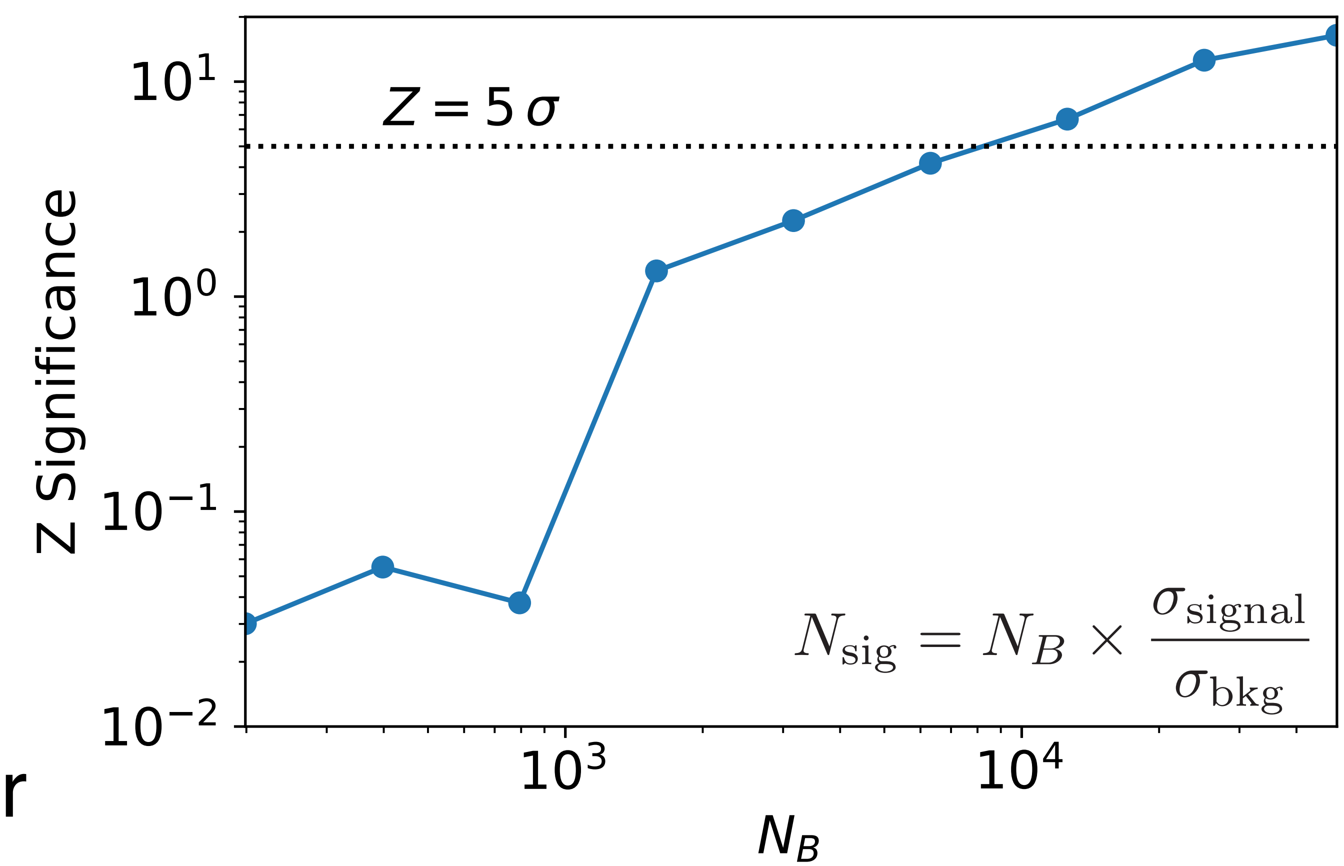
Expt. Collab. at CERN interested in applying this test

> DM search @ LHC



stronger signal
easier to discover

$$N_T = N_B + N_{\mathrm{sig}}$$



more data,
more power

> Outlook

Directions for future work:

- inclusion and impact of uncertainties

- adaptive choice of *K*

- identifying discrepant regions in realistic situations
  (with *Z*-score method)

- validation tool for bkg:
  compatibility between MC sims. and data in control regions

- scalability

> Take-Home Messages

1. Golden age of Machine Learning

Big Data are everywhere

2. Innovative Statistical Test for New Physics

- Powerful and model-independent discovery tool
- Guidance for experimental searches

3. ML for science in germinal stage

Pioneering developments waiting ahead!