Carleton
Feb 1, 2018

On the Topic of Jets

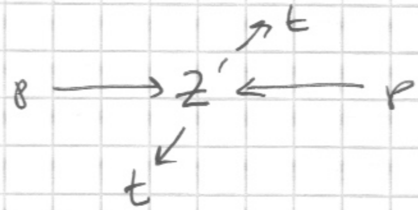
1802.0xxxx

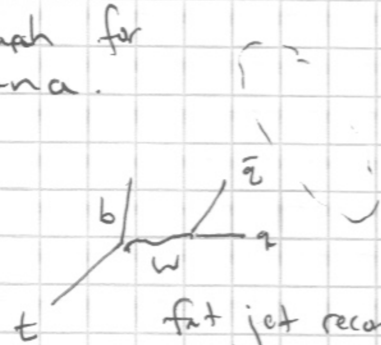
JDT, Eric Metodiev, prior work with Ben Nachman

I've been working on jets for roughly a decade.

- Initial goal: Enhance the search for new phenomena.

e.g.



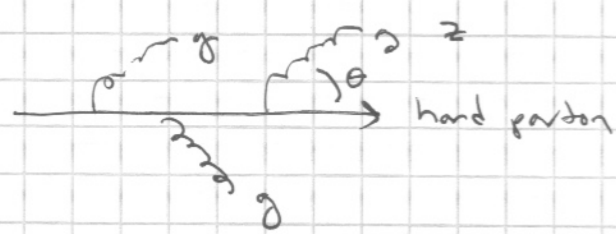


fat jet reconstruction

$$\Delta R \sim \frac{1}{\gamma_{top}}$$

- Current goal: Expose underlying structures of QCD (and other QFTs)

e.g.



$$P_{i \to ig} \simeq \frac{2\alpha_s}{\pi} C_i \frac{dz}{z} \frac{d\theta}{\theta}$$

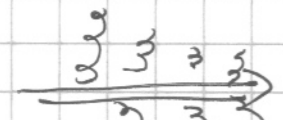
quarks = 4/3
gluons = 3

Can you measure α_s from substructure of jet?

Today's talk grew out thinking about recent progress on machine learning.

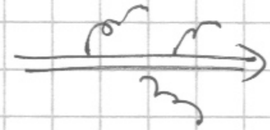
Underlying categories.

g



C_A vs. C_F in radiation pattern

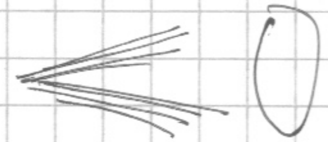
u d s



c b

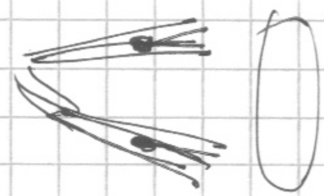
displaced vertex

boosted W/Z



two prongs at 80/91 GeV

boosted H



two b-tagged prongs at 125 GeV

boosted top

three prongs, one b-tagged at 170 GeV

parton-level objects (sometimes colored) $\Longrightarrow$ hadron-level reconstruction (color singlet)

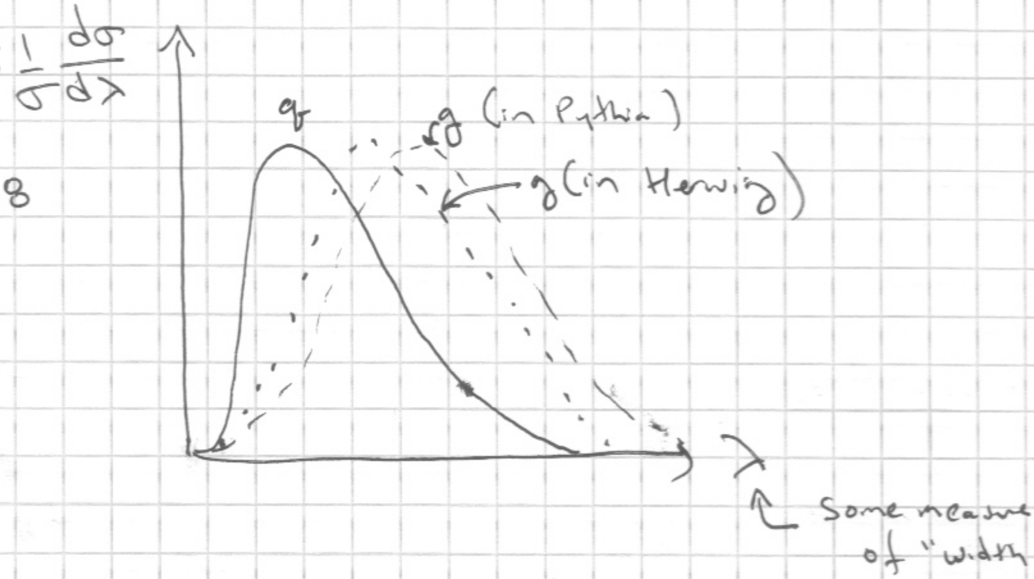
Map is fundamentally ambiguous, but how much discrimination power can you achieve?

E.g. Quark / Gluon Discrimination. (almost as old as QCD!) ③

Answer 1: Train your favorite ML algorithm on parton shower MC generators (i.e. Pythia, Herwig, Sherpa.)

Problem 1:

1704.03878 (systematics of quark/gluon tagging)



Smallish uncertainties on distributions change story from "always q/g tag" to "no point to q/g tag".

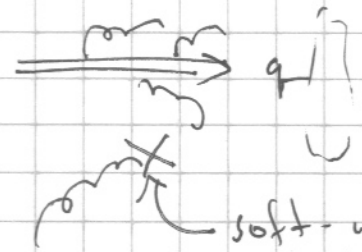
Answer 2: Train your favorite ML algorithm directly on data.

Problem 2: Need to know quark/gluon fractions!

Problem 2': Need to know that "quark" in one sample is same as "quark" in another.

$\Longrightarrow$ "Sample Dependence"

Answer 3′: Use "jet grooming" algorithms to mitigate sample dependence.



soft-wide angle color correlations, can be systematically removed.

Answer 3: In many cases, you can learn quark/gluon fractions from the data itself. (!)

Key Assumption:

$$p_a(\vec{\lambda}) = f_q^{(a)} \, p_q(\vec{\lambda}) + (1 - f_q^{(a)}) \, p_g(\vec{\lambda})$$

various mixed samples (for a); set of jet properties (for $\vec{\lambda}$)

There are robust analysis techniques to simultaneously extract $f_q^{(a)}$, $p_q(\vec{\lambda})$, $p_g(\vec{\lambda})$.

Almost like "learning from nothing"
(except assumption is rather strong to begin with)

Has nothing really to do with "machine learning".

Actually just follows from saying that $p_{\vec{v},\vec{g}}(\vec{\lambda}) > 0$

(because it is a probability)

In retrospect, quite obvious. But instructive

to explain the way we got there.

Classification without labels.

I have a bunch of cat videos

I have a bunch of dog videos

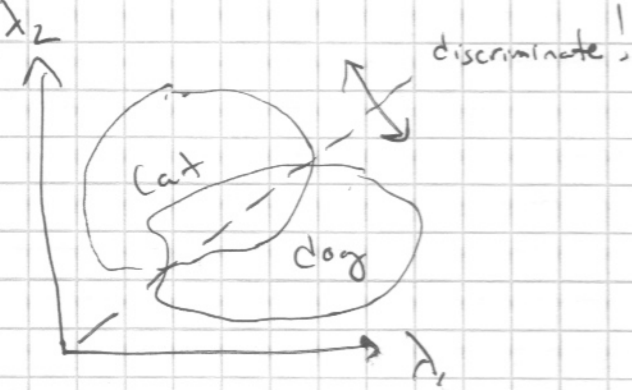
I want (or google wants) to classify

videos as cat-like or dog-like.

$p_{cat}(\vec{\lambda})$

features of ~~...~~ videos

$p_{dog}(\vec{\lambda})$



If you knew $p_{cat}(\vec{\lambda})$ and $p_{dog}(\vec{\lambda})$ exactly,
then "Neyman - Pearson lemma" states that
best you can do is select videos with

(in practice, ML algorithms try to approximate likelihood ratio)

$$\frac{p_{cat}(\vec{\lambda})}{p_{dog}(\vec{\lambda})} > \text{Cut value.}$$

Best cat efficiency per dog rejection.

Now, imagine you dropped (digitally) your cat/dog video folders, and you get two mixtures.

$$p_{Mix\,A}(\vec{\lambda}) = f^{A}_{cat}\, p_{cat}(\vec{\lambda}) + (1 - f^{A}_{cat})\, p_{dog}(\vec{\lambda})$$
$$p_{Mix\,B}(\vec{\lambda}) = f^{B}_{cat}\, p_{cat}(\vec{\lambda}) + (1 - f^{B}_{cat})\, p_{dog}(\vec{\lambda})$$

Seems like you are out of luck! Best you can do is separate mixture A from mixture B.

$$\frac{p_A(\vec{\lambda})}{p_B(\vec{\lambda})} > \text{cut value}$$

But is this at all useful to tell cat from dog?!

Remarkably, yes!

$$\frac{p_A}{p_B} = \frac{f_A\, p_{cat} + (1-f_A)\, p_{dog}}{f_B\, p_{cat} + (1-f_B)\, p_{dog}}$$

$$= \frac{f_A \frac{p_{cat}}{p_{dog}} + (1-f_A)}{f_B \underbrace{\frac{p_{cat}}{p_{dog}}}_{\text{what I want!}} + (1-f_B)}$$

The trick: (monotonicity)

$$\frac{\partial(p_A/p_A)}{\partial(p_{cat}/p_{dog})} = \frac{f_A - f_B}{\left(1 + f_B\left(\frac{p_{cat}}{p_{dog}} - 1\right)\right)^2} > 0$$

as long as $f_A > f_B$.

So a cut on a "mix A" / "mix B" classifier has the same effect as a cut on the optimal dog/cat classifier.

$\Longrightarrow$ CWoLa. (JDT, Metodiev, Nachman) 1708.02949.

As long as you have two mixtures, and you know $f_A > f_B$, then you can train towards optimal algorithm.

Too good to be true?

Yes. You get optimal discriminant, but you don't know cat efficiency / dog rejection without a separate calibration step.

Email from Mario Campanelli (ATLAS):

Can't you use CWoLa to find optimal separation variable, do a measurement on LHC data applying some cut, and let theorists N years from now figure out calibration after the fact?

(He has a lot of faith in theorists ability to calculate in the large N limit...)

In parallel, with Eric Metodiev

Would you even know what quark/gluon success looks like? No unambiguous quark/gluon definitions, so is there even a calculation in theory that could work?

And thus Eric stumbles into world of

"Topic Modelling" (and many other related fields, some already used for CMB physics)

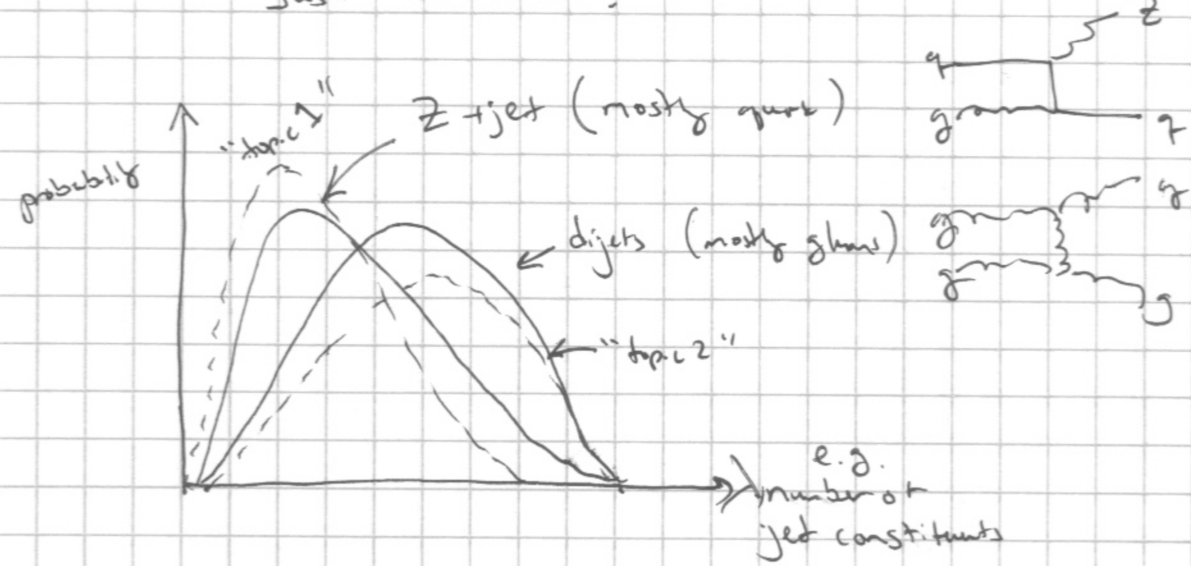
You have a	Corpus	Collection of Histograms
consisting of a bunch of	Documents	Histograms of Jet Observables
You want to know if they are about a particular (without having to specify the topic ahead of time)	Topic	Type of Jet ("jet topic") → (!)
You know that there is a common	Vocabulary	Jet Observable
And documents select from this specific (unordered) ← Strange assumption to make for word documents...	Words.	Histogram bin.

How well can you do?

Entire industry trying to "demix" mixed samples. (including the Demix algorithm)
1710.01167
Katz-Samuels, Blanchard, Scott

If you have just two samples and two types, you don't need a fancy algorithm...

Just subtract!



$$P_{\text{topic }1}(\vec{\lambda}) = \frac{P_{Z+\text{jet}}(\vec{\lambda}) - K\, P_{\text{dijets}}(\vec{\lambda})}{1-K}$$

Choose K as big as possible such that $P_{\text{topic}1}(\vec{\lambda})$ is everywhere positive.

Do the same for topic 2.

Declare victory. By construction, answer is independent of initial fractions, and can be inverted to find fractions.

Key Assumptions: Sample Independence
Different Mixtures.
Same as CWoLa

"Mutual Irreducibility"
↳ there is some region of phase space that is only quark
(and another that is only gluon)

In the topic modeling literature, need to have
"anchor word" = Word that only ever appears in one topic.

energy

Kyoto protocol

anchor to climate science.

conservation

ambiguous.

Noether's theorem

anchor to physics

Too good to be true?

Yes... Can show that some jet substructure observers do not yield "anchor" bins.
(though you can in principle correct using analytic calculations

On the other hand...

This is no unambiguous definition of quark/gluon jets, so maybe you want to <u>define</u> "quark" and "gluon" in terms of "mutually irreducible topics"

But in practice, go back to Mario's point.
As long as you can apply the same algorithm of data and (eventually) on theory, you are all set.

Many, many applications where q/g fraction information is needed
- PDF extraction
- Monojet searches for dark matter
- α_s extraction (b/c. $C_A \alpha_s$ vs. $C_F \alpha_s$ is huge uncertainty)
- quark/gluon jet quenching in heavy ions.

Lesson: If there are underlying categories in data, you can just find them (even if you don't know what they are!).
New opportunity for data-driven science