iv) point clouds

In fact jet const. have no canonical/natural order

role of symmetry in ML architecture design

top tagger should be permutation inv't!

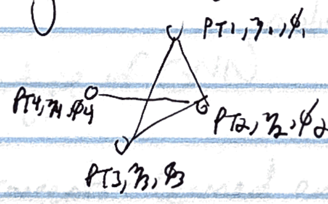
$$p(J=(\vec{p}_1,\vec{p}_2,\dots\vec{p}_N)) = p(J'=(\vec{p}_2,\vec{p}_1,\dots\vec{p}_N))$$

points to better representation: <u>point cloud: set or unordered list</u>

ML architectures: Graph NNs

Transformers

"Message passing": each jet const. a node in a graph



$\forall$ edge

$\phi^{edge}(e,a,b)=e'$

ϕ^{edge}, ϕ^{node} small MLPs

weight sharing!

$\forall$ node

$\phi^{node}(\rho(e_1',e_2',\dots),a) = a'$

<u>perm inv'ce</u> $\leftarrow$ ρ = symmetric fn

produces another graph $\rightarrow$ stackable!

<u>Transformer:</u> Workhorse of LLM, AI, NLP

Probably single most important modern ML architecture!

— Transformers key to modeling language — sequences

However! They are <u>permutation equivariant</u>

So actually point clouds!

$(t_1, t_2, \dots) \longrightarrow \text{Transf.} \longrightarrow (t_1', t_2', \dots)$

So they can learn highly nonlocal correlations/patterns

major advance in NLP!

— Can think of as a kind of input-dependent MLP

$X \to WX$ MLP W fixed weights

$X \to W(X)X$ Transf W input-dep.

nonlinear → much more expressive!

— Also a type of GNN

— Key concepts: learned embeddings & attention mechanism

Map sequence/point cloud → embedding space

$\vec{t}_1, \vec{t}_2, \dots \vec{t}_N \xrightarrow{\text{embedding}} \vec{v}_1, \vec{v}_2, \dots \vec{v}_N$

Relations btw elements of sequence

expressed by simple dot product in embedding space

$\vec{v}_i \cdot \vec{v}_j$

"Attention matrix" $N \times N$

NLP example: "the animal didn't cross the street because it was tired"

What does "it" refer to? animal? street?

attention ~1 for animal

~0 for street.

Attention Mechanism

Data : $X = \begin{pmatrix} x_1 — \\ x_2 — \\ \vdots \\ x_N — \end{pmatrix}$ $N \times d$ matrix

weights W^Q $d \times d'$ (indep of seq location — weights stay!)

W^K $d \times d'$

W^V $d \times d'$

$Q = XW^Q$ $N \times d'$ query

$K = XW^K$ $N \times d'$ key

$V = XW^V$ $N \times d'$ value

- self-atten mtx $P = a(Q . K^T)$ $N \times N$ mtx

↳ "softmax" turns rows into probs

$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ \vdots & & & \\ p_{N1} & - & - & p_{NN} \end{pmatrix}$ $\sum_{j=1}^{N} p_{ij} = 1$, $1 \geq p_{ij} \geq 0$.

- Transformer output

$X \to X' = P V$ $N \times d'$ matrix

seq $\underset{d \text{ dim}}{x_1, \cdots x_N} \longrightarrow \underset{d' \text{ dim}}{x_1', \cdots x_N'}$

— it is seq → seq mapping, x_i' depends on all of $x_1, \cdots x_N$

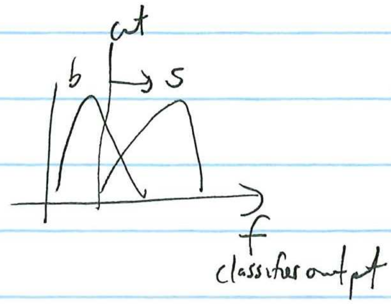
— it is perm-equivariant $x_i \leftrightarrow x_j \longrightarrow x_i' \leftrightarrow x_j'$.

— Can stack!

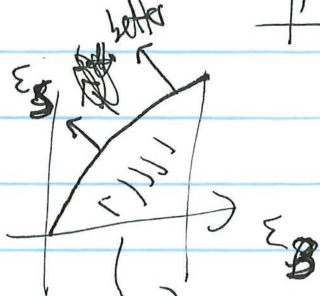
— elaborations — "multiheaded attn", encoder/decoder, ...

Metrics for classification

- loss ~~[scribbled out]~~
- ROC curve



$f > f_c \longrightarrow$ ε_s "tpr" sig eff, ε_b "fpr" bg

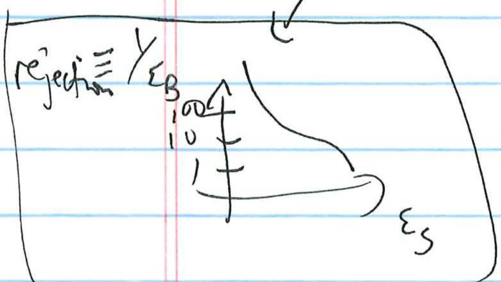


"AUC"

$$\max\left(\frac{\varepsilon_s + 1 - \varepsilon_b}{2}\right) \doteq \text{"ACC"}$$

AUC = ACC = 0.5 "random"

~~A~~ 1 1 "perfect"

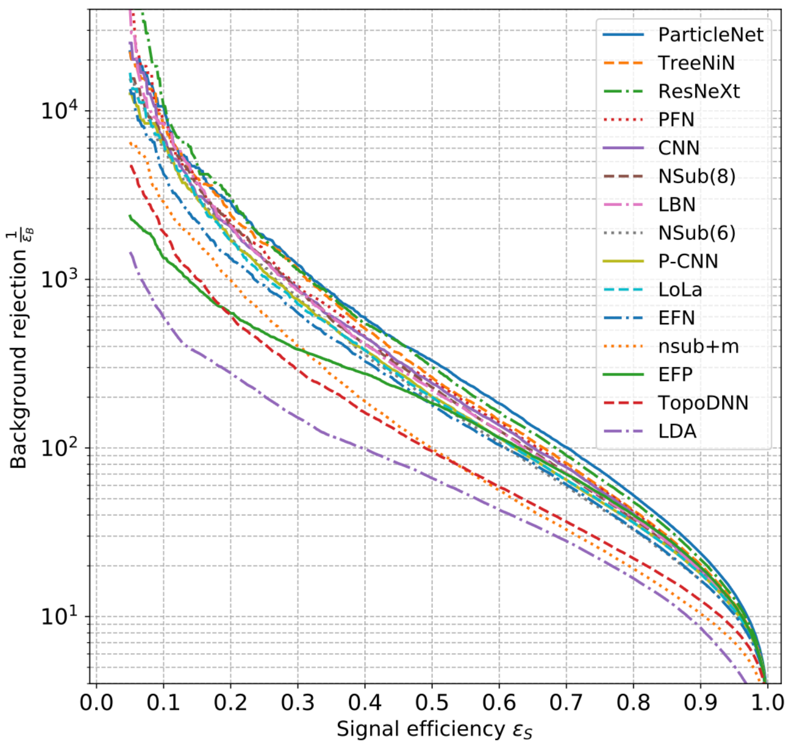


more common in our field

→ R30 : bg rej @ $\varepsilon_s = 0.3$ a useful metric!
(more sensitive than acc/auc)

$\boxed{\text{DEMO}}$

Top tagging



"The Machine Learning Landscape of Top Taggers"
Kasieczka, Plehn (ed) et al 1902.09914

Table 5. Comparison between ParT and existing models on the top quark tagging dataset. ParT refers to the model trained from scratch on this dataset. ParticleNet-f.t. and ParT-f.t. denote the corresponding models pre-trained on JetClass and fine-tuned on this dataset. Results for other models are quoted from their published results: P-CNN and ParticleNet (Qu & Gouskos, 2020), PFN (Komiske et al., 2019b), JEDI-net (Moreno et al., 2020), PCT (Mikuni & Canelli, 2021), LGN (Bogatskiy et al., 2020), rPCN (Shimmin, 2021), and LorentzNet (Gong et al., 2022).

	Accuracy	AUC	$\mathrm{Rej}_{50\%}$	$\mathrm{Rej}_{30\%}$
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	—	0.9819	247 ± 3	888 ± 17
ParticleNet	0.940	0.9858	397 ± 7	1615 ± 93
JEDI-net (w/ $\sum O$)	0.930	0.9807	—	774.6
PCT	0.940	0.9855	392 ± 7	1533 ± 101
LGN	0.929	0.964	—	435 ± 95
rPCN	—	0.9845	364 ± 9	1642 ± 93
LorentzNet	0.942	0.9868	498 ± 18	2195 ± 173
ParT	0.940	0.9858	413 ± 16	1602 ± 81
ParticleNet-f.t.	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t.	**0.944**	**0.9877**	$\mathbf{691 \pm 15}$	$\mathbf{2766 \pm 130}$

"Particle Transformer for Jet Tagging"
Qu, Li, Qian 2202.03772