

Esplorazione di metodi avanzati di auto-apprendimento statistico per la classificazione ottimale di collisioni finalizzata alla fisica del sapore

Candidato: Marco Sciorilli

Relatore: Diego Tonelli

Correlatore: Eldar Ganiev



Dipartimento di Fisica
Università degli Studi di Trieste

19 ottobre 2020

Fisica del sapore per superare il Modello Standard

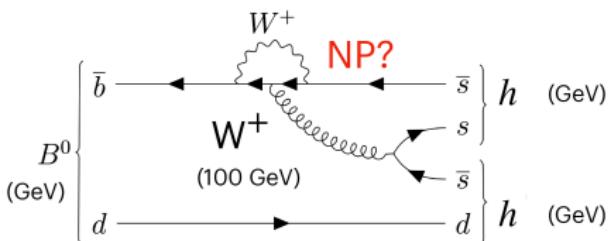
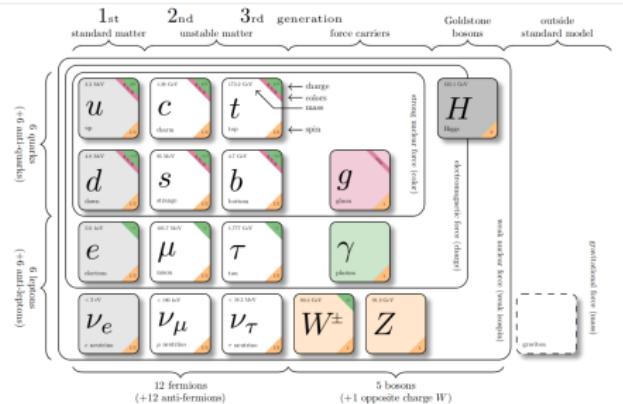
Modello Standard (SM): predizioni accurate dall'eV al TeV con soli 20 parametri.

Domande aperte

- ▶ Materia oscura?
 - ▶ Asimmetria materia-antimateria?
 - ▶ Discrepanza intensità forza gravitazionale e altre forze?
 - ▶ ...

Estensione del SM a $E > 10$ TeV è obiettivo primario della fisica delle particelle.

Fisica del sapore testa i limiti dello SM: confronto tra misure e predizioni ad alta precisione sensibile a contributi di particelle non-SM virtuali di massa fino a 1000 TeV.



L'esperimento Belle II a superKEKB

SuperKEKB è un collisore e^+e^- a 10 GeV, la soglia di produzione della risonanza $\Upsilon(4S)$, che accade in 0.4 % delle collisioni e decade in $B\bar{B}$ (sia B^+B^- che $B^0\bar{B}^0$).

Rivelatore Belle II:

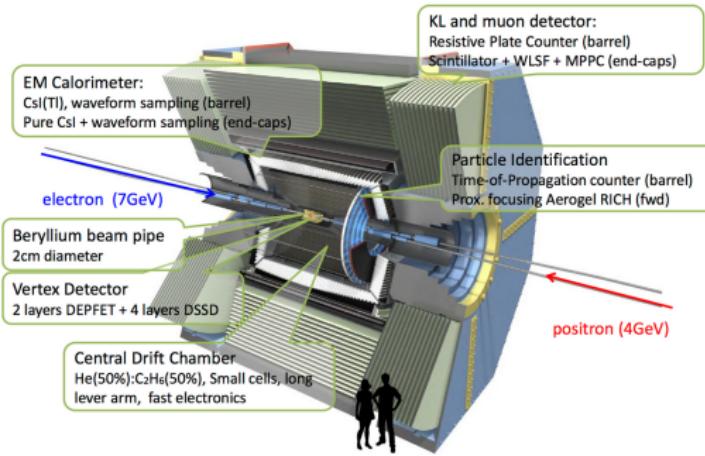
- ▶ Rivelatore di vertice al silicio
 - ▶ Camera a multifili centrale
 - ▶ Rivelatore di identificazione di particelle
 - ▶ Calorimetro elettromagnetico
 - ▶ Rivelatori per K^0 e μ

Verranno prodotte ad intensità nominale 1000 coppie $B\bar{B}$ al secondo.

Ad oggi ha raccolto 80M di coppie $B\bar{B}$

Primi dati essenziali per comprendere ed ottimizzare le prestazioni del rivelatore e della ricostruzione

Belle II Detector



Obiettivo: ottimizzazione della selezione del canale di decadimento

$$B^0 \rightarrow D^- [\rightarrow K^+ \pi^- \pi^-] \pi^+$$

con un metodo di auto-apprendimento statistico, il *boosted decision-tree*.

Importante per somiglianza a $B^0 \rightarrow D^- [\rightarrow K^+ \pi^- \pi^-] K^+$ fondamentale per misure di violazione simmetria carica-parità, tra gli obiettivi primari di Belle II.

- ▶ Strumenti:
 - Simulazione Monte Carlo.
 - 5 milioni di coppie $B\bar{B}$ da primissimi dati Belle II (2019).
- ▶ Analisi e scelta delle variabili discriminanti.
- ▶ Implementazione del *decision-tree*.
- ▶ Studio delle prestazioni del classificatore.
- ▶ Analisi della classificazione e risultati.

Primo contatto coi dati

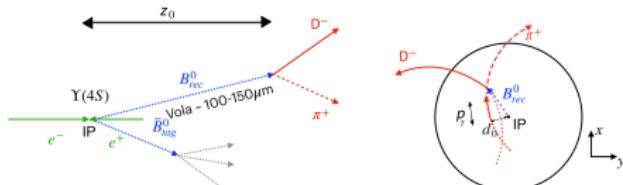
$$BF(B^0 \rightarrow D^- \pi^+) \times BF(D^- \rightarrow K^+ \pi^- \pi^-) \approx 2 \times 10^{-4}$$

B^0 : $m \approx 5280 \text{ MeV}/c^2$, $\tau \approx 1.5 \text{ ps}$

D^- : $m \approx 1870 \text{ MeV}/c^2$, $\tau \approx 1 \text{ ps}$

K^+ : $m \approx 494 \text{ MeV}/c^2$, stabile nel detector

π^+ : $m \approx 140 \text{ MeV}/c^2$, stabile nel detector



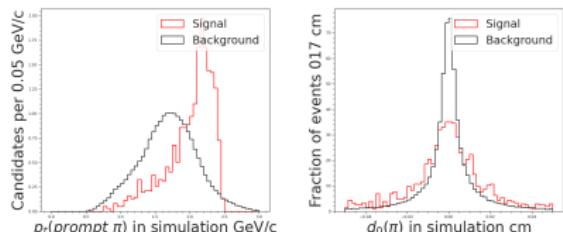
Simulazione

- ▶ distanza trasversale della traiettoria dal punto di collisione (d_0),
- ▶ distanza longitudinale (z_0),
- ▶ componente dell'impulso trasversa al fascio nel sistema del centro di massa (p_t),

misurati per particelle finali.

L'energia del B si distribuisce tra un numero inferiore di particelle che nel fondo: p_T maggiori.

B decade dopo aver percorso 150 μm : la traiettoria non punta al punto di collisione, $d_0 \neq 0$ più probabile .



Punto di partenza

Selezione preliminare

- ▶ Ridurre tracce da interazioni fascio
 - $|d_0| < 0.5$ cm
 - $|z_0| < 3$ cm
- ▶ Ridurre fondo di π
 - kaonID (soglia su kaoni) > 0.6
- ▶ Massa invariante candidati $K^+ \pi^- \pi^-$ vincolata ad essere compatibile con quella del mesone D
- ▶ $-0.15 < \Delta E < 0.15$ GeV

Misuriamo segnale con ΔE : differenza tra l'energia osservata del candidato B e metà dell'energia di collisione.

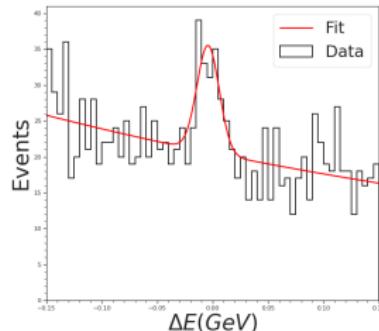
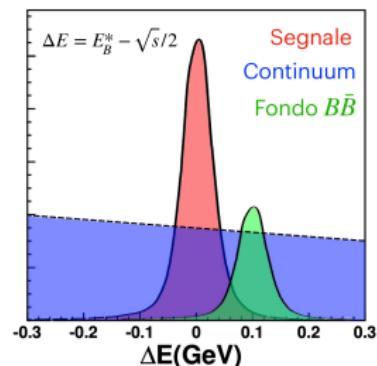
$\Upsilon(4S)$ decade in $B\bar{B}$.

- ▶ Se è presente un **B ricostruito correttamente**, la sua energia è metà dell'energia di collisione.
⇒ picco a $\Delta E = 0$
- ▶ Per fondo o **B ricostruito erroneamente**, l'energia del candidato B differisce dall'energia vera, quindi $\Delta E \neq 0$.

Fit di χ^2 : esponenziale decrescente + Gaussiana.

75 ± 16 decadimenti di segnale.

Rapporto segnale-fondo al picco di 0.6.



Decision trees

Strumento di auto-apprendimento statistico.
efficace per classificazione in problemi non lineari, come selezioni in fisica dei collider.

"Impara" le soglie di criteri binari su variabili discriminanti scelte minimizzando una funzione di costo su un campione (tipicamente simulato) a classificazione nota.

Applicato sui dati, classifica come segnale o fondo in base alla classificazione precedente.

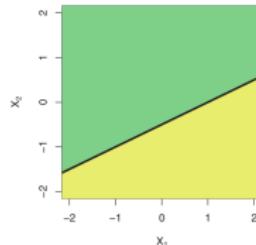
Tende a *data-dependence*: "impara" fluttuazioni del campione di training, perdendo generalità.

Boosting

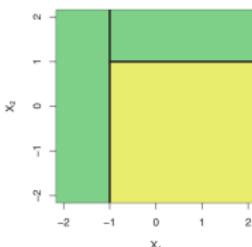
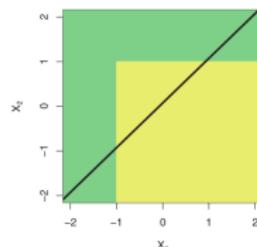
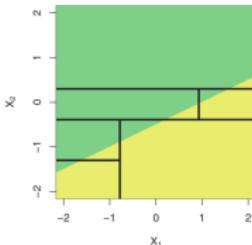
- ▶ Creare multipli *decision-tree*
- ▶ Media pesata sui tree per sopprimere fluttuazioni.

Scelta delle variabili discriminanti è cruciale per l'ottimizzazione del classificatore.

Separazione lineare



Decision-tree



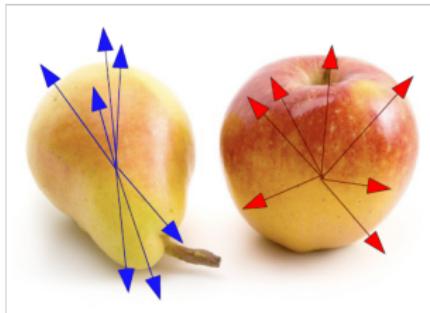
Scelta delle variabili discriminanti

Voglio discriminare

- ▶ **Fondo da quark leggeri**, 10^5 volte più frequenti del segnale, massa $\ll 10$ GeV, prodotti in getti collimati back-to-back con alti impulsi.
- ▶ **Segnale da coppie mesoni $B\bar{B}$** , massa ≈ 10 GeV, praticamente ferme, decadono isotropicamente.

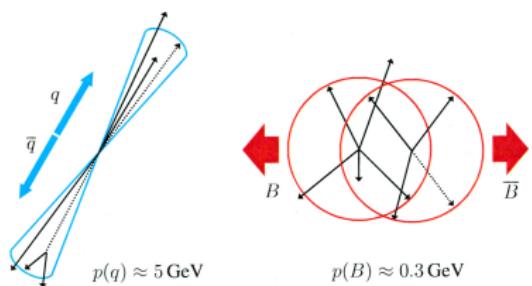
Devo:

- ▶ Sfruttare **differenze** di "forma" dei flussi di energia-impulso dell'evento tra segnale e fondo.
- ▶ Scegliere variabili **ben modellate dalla simulazione**, poiché il classificatore addestrato sulla simulazione sia efficace anche sui dati.



Quark leggeri

$B\bar{B}$



$$e^+e^- \rightarrow q\bar{q} \quad (q \in \{u, d, s, c\})$$

$$e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$$

La *spinta*:

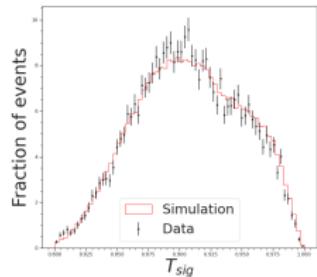
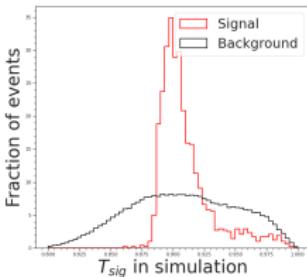
$$T = \frac{\sum_{i=1}^N |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_{i=1}^N |\mathbf{p}_i|}$$

Varianti:

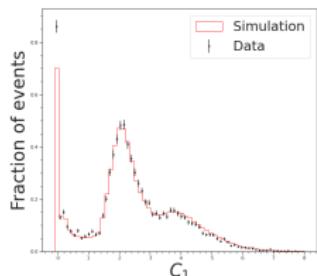
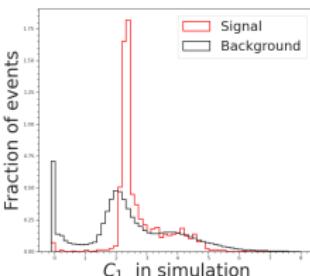
- ▶ Angolo di *spinta*: angolo tra spinta di candidato B e spinta del resto dell'evento.
- ▶ CLEO cones: somma impulsi compresi in 9 aree coniche concentriche.

Tutte queste sono buoni discriminanti.

Spinta



CLEO cone 1



Netta distinzione tra segnale e fondo in simulazione.

Dati ben rappresentati dalla simulazione.

I momenti Fox-Wolfram:

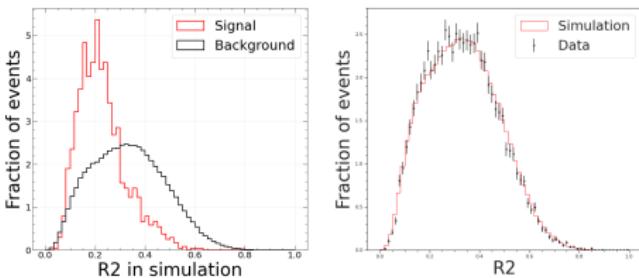
$$H_I = \sum_{i,j}^N \frac{|\mathbf{p}_i^*| \cdot |\mathbf{p}_j^*|}{s} \cdot P_I(\cos\theta_{i,j}^*),$$

Varianti:

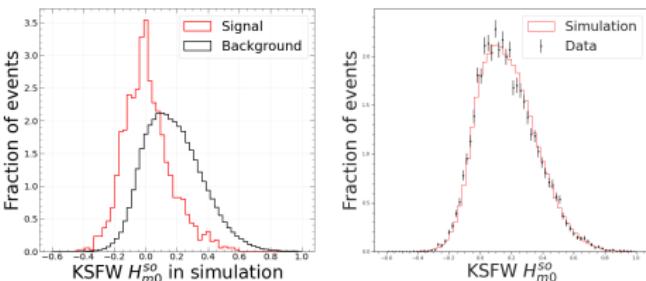
- ▶ Momenti Kakuno-Super-Fox-Wolfram tengono conto delle particelle non rilevate.

Buoni discriminanti.

Secondo momento FW



Primo momento KSFW



Netta distinzione tra segnale e fondo in simulazione.

Dati ben rappresentati dalla simulazione.

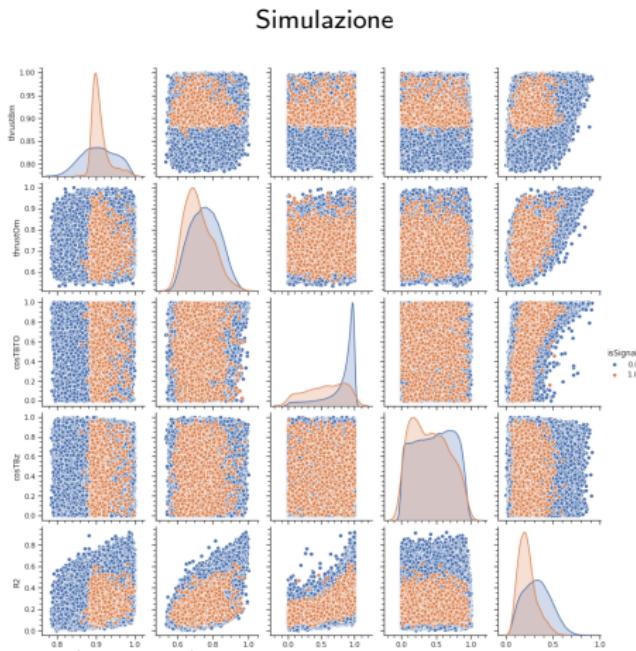
Pruning delle variabili

Potenza di calcolo finita.

Necessaria scelta delle variabili.

Scartare quelle con minor potere discriminante.

Grafici di correlazione identificano variabili più discriminanti.



Spinta efficace in tutti gli abbinamenti.

Forte correlazione tra gli angoli di *spinta*.

Prestazioni del classificatore

ROC = reiezione del fondo in funzione di efficienza di segnale.

Area sotto ROC: probabilità classificatore identifichi correttamente l'evento.

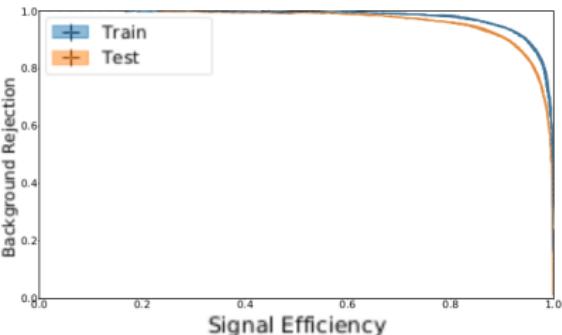
Data set simulato diviso in due: addestrato il classificatore sul primo, lo testo sul secondo.

Minima differenza tra ROC curves: il classificatore non è *data-dependent*.

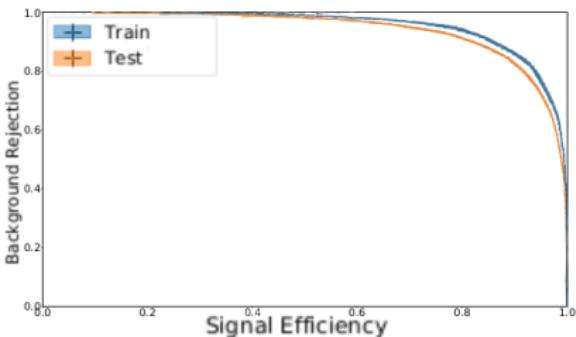
Restringendomi alle 11 variabili più discriminanti, semplifico calcoli con marginale degradazione di prestazioni.

Scelgo classificatore con migliori prestazioni, usando tutte 27 variabili studiate.

Simulazione - 27 variabili

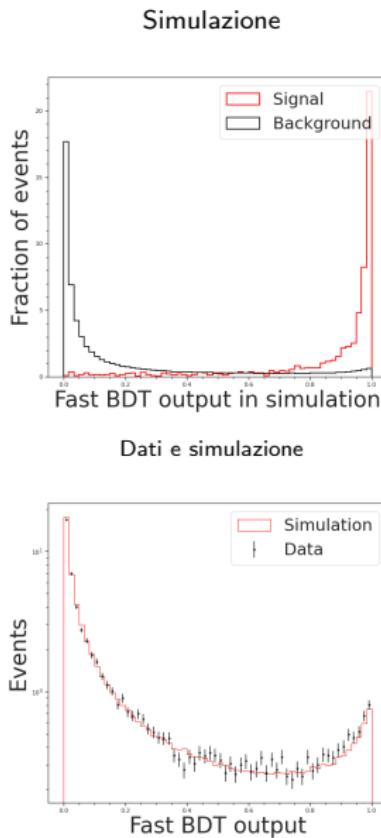


Simulazione - 11 variabili



Output del classificatore

Uno scalare tra 0 ed 1 per ogni evento.
Distribuzioni di output tra simulazioni e dati coincidono.
Il *boosted decision-tree* è un buon modello delle caratteristiche dei dati.

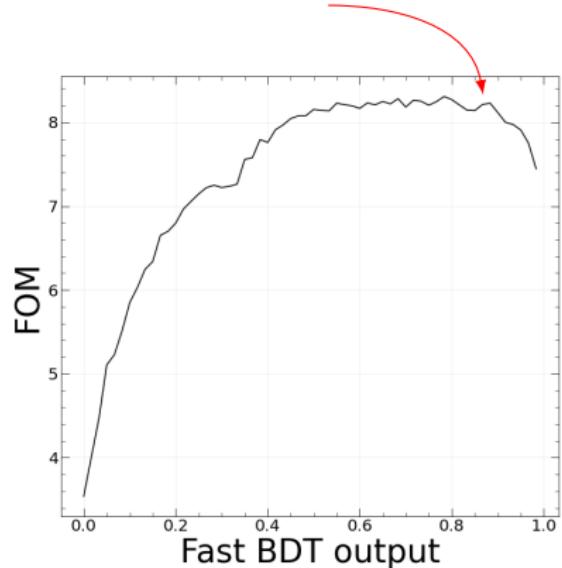


Scelgo soglia massimizzando

$$FOM = \frac{N_S}{\sqrt{N_S + N_B}}$$

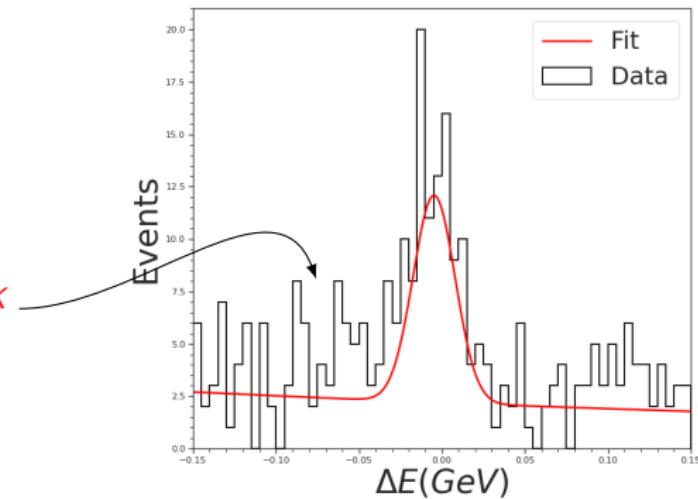
N_S (N_B) numero di eventi osservati di segnale (fondo).

Il massimo corrisponde a $FBDT > 0.90$.



Risultati

- ▶ Riduzione significativa fondo.
- ▶ 63 ± 7 eventi di segnale.
- ▶ Segnale-fondo al picco 2.5.
- ▶ Osservo un'indicazione di $B^0 \rightarrow DK$ in $-0.1 < \Delta E < -0.05$ GeV.



Fisica del sapore è centrale per cercare estensioni del Modello Standard.

Belle II avrà un ruolo di primo piano.

Ho sviluppato ed ottimizzato un classificatore per isolare decadimenti importanti nei primissimi dati dell'esperimento.

Use $B^0 \rightarrow D^- \pi^+$ come benchmark.

- ▶ Migliorato rapporto segnale-fondo di fattore 4.
 - ▶ Conservato 84% del segnale.

Sostanziale miglioramento rispetto ai metodi convenzionali basati su tagli.

