

Identification of low-energy antiprotons on pi-meson background using machine learning methods in PAMELA experiment

A. Lukyanov¹, A. Mayorov^{1,2}, O. Dunaeva¹, Yu. Bogomolov¹ et al.
on behalf of PAMELA collaboration

¹ Yaroslavl State University ² National Research Nuclear University MEPHI

2016

Classification of antiprotons and pi-mesons using track system features

- ▶ Dataset is classified using Support Vector Machines (SVM) classifier with RBF kernel.
- ▶ Dataset size: 300000 events.
- ▶ Each object in dataset is described by the following 16 features

Track system features

$$R_m$$

Measured rigidity.

$$\left\langle \frac{dE}{dx} \right\rangle_{12}, \left\langle \frac{dE}{dx} \right\rangle_{6X}, \left\langle \frac{dE}{dx} \right\rangle_{6Y}$$

Average ionization loss.

$$\frac{dE}{dx} x(1), \frac{dE}{dx} y(1)$$

Independent measurements of ionization loss of the first plane in x and y projections.

...

$$\frac{dE}{dx} x(6), \frac{dE}{dx} y(6)$$

Independent measurements of ionization loss of the sixth plane in x and y projections.

Classification using standard criterion

- ▶ $\frac{dE}{dx} = 10.8 \cdot \exp(-2.95 \cdot (\text{abs}(R) + 0.12)) + 40.7 \cdot \exp(-11.8 \cdot (\text{abs}(R))) + 0.8$
- ▶ $\frac{dE}{dx}$ — average rigidity over all planes, R — rigidity.
- ▶ If event's $\frac{dE_i}{dx}$ value is higher than $\frac{dE}{dx}$ then this is an antiproton, otherwise this is a pi-meson.

Simple classification based on average value of ionization loss in track system show the following results (average accuracy 71%):

| predicted \ true | pi-mesons | antiprotons |
|------------------|-----------|-------------|
| pi-mesons | 39% | 61% |
| antiprotons | 0.119% | 99.9% |

Standard criterion accuracy

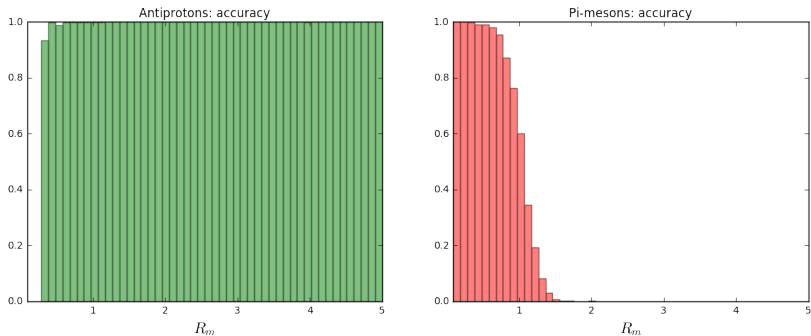


Figure: Fraction of correctly classified events according to measured rigidity R_m .

Classification using SVM

However, our goal is to correctly classify as many antiproton events as possible and at the same time provide good level of correct classification for pi-mesons. For this we configure SVM so that it provides unbalanced classification.

Classification using SVM

Standard SVM shows average accuracy of 80%:

| | predicted | | |
|-------------|-----------|-----------|-------------|
| true | | pi-mesons | antiprotons |
| pi-mesons | | 66% | 34% |
| antiprotons | | 8% | 92% |

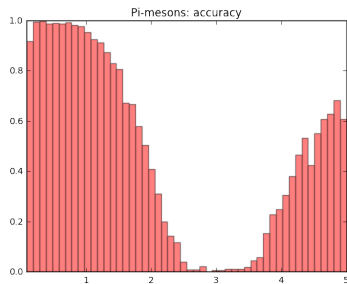
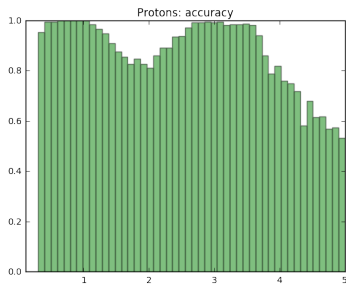


Figure: Standard SVM accuracy.

Classification using SVM

$w = 1.1$, accuracy 80%:

| true \ predicted | pi-mesons | antiprotons |
|------------------|-----------|-------------|
| pi-mesons | 63.7% | 36.3% |
| antiprotons | 5.7% | 94.3% |

$w = 1.4$, accuracy 79%:

| true \ predicted | pi-mesons | antiprotons |
|------------------|-----------|-------------|
| pi-mesons | 59.1% | 40.9% |
| antiprotons | 2.8% | 97.2% |

$w = 3$, accuracy 77%:

| true \ predicted | pi-mesons | antiprotons |
|------------------|-----------|-------------|
| pi-mesons | 51.8% | 48.2% |
| antiprotons | 0.5% | 99.5% |

Classification using SVM

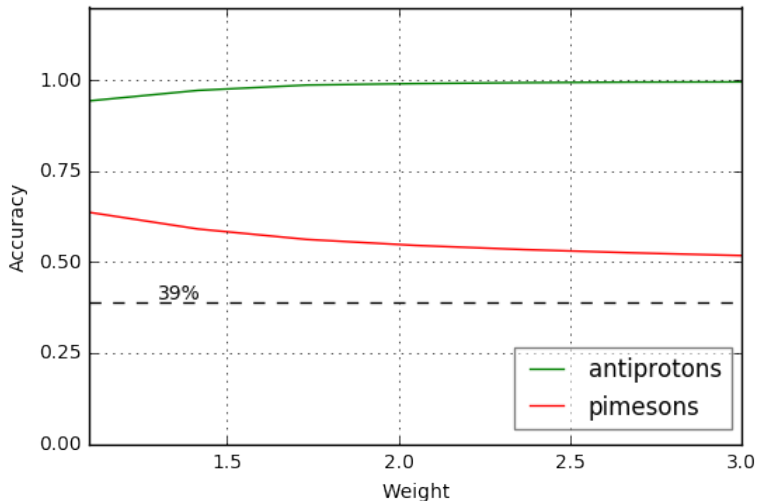
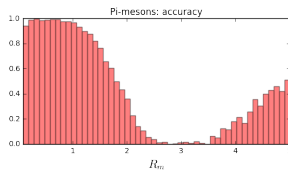
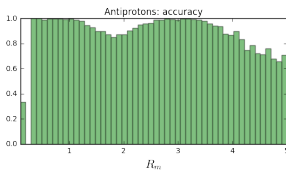


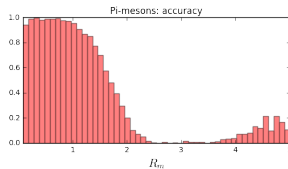
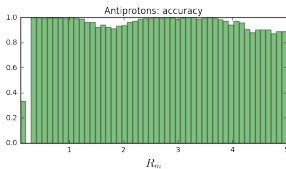
Figure: Classification accuracy of antiprotons and pi-mesons according to increasing class weight for antiprotons.

Classification using SVM

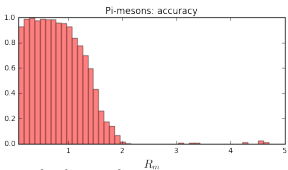
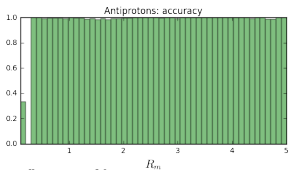
$w = 1.1$, accuracy 80%:



$w = 1.4$, accuracy 79%:

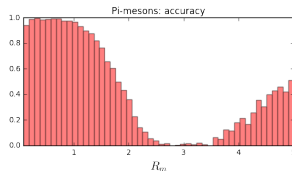
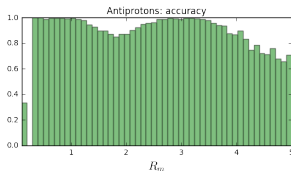


$w = 3$, accuracy 77%:

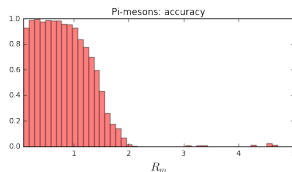
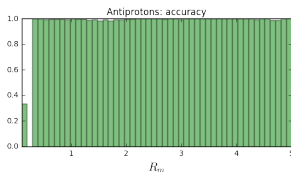


Compared with standard criterion

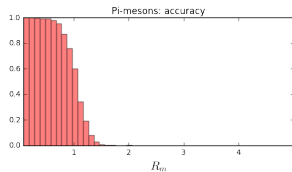
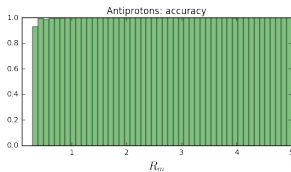
Balanced SVM (accuracy 80%):



Unbalanced SVM (accuracy 79%):



Standard criterion (accuracy 71%):



Classification using track system and calorimeter features

In order to improve classification quality after 2 GV we add the 17 calorimeter features to the descriptor:

Q_{core} Total energy release in both projections in the cylinder of radius 8 strips.

Q_{tot} Total energy release in the calorimeter.

Q_{max} Maximum energy release in the calorimeter.

Q_{track} Energy release along the shower axis.

Q_{tr} Energy release in the cylinder with radius of 4 strips around the shower axis.

Q_{cyl} Energy release in the cylinder with radius of 8 strips around the shower axis.

N_{cyl} The number of triggered strips in the cylinder with radius of 8 strips around the axis of the shower.

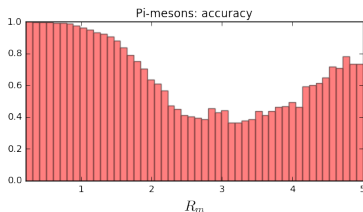
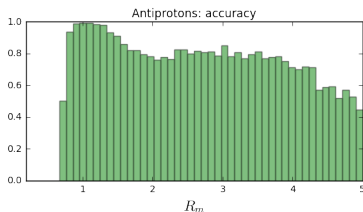
Q_{pre} The energy release in the cylinder radius of 8 strips around the axis of the shower in the first three planes.

...

Classification results

Balanced SVM shows the following results (average accuracy 78%):

| true \ predicted | predicted | |
|------------------|-----------|-------------|
| | pi-mesons | antiprotons |
| pi-mesons | 73.2% | 26.8% |
| antiprotons | 16.6% | 83.4% |



Tuning SVM

$w = 1.1$, accuracy 78%:

| true \ predicted | pi-mesons | antiprotons |
|------------------|-----------|-------------|
| pi-mesons | 70% | 30% |
| antiprotons | 14% | 86% |

$w = 1.9$, accuracy 75%:

| true \ predicted | pi-mesons | antiprotons |
|------------------|-----------|-------------|
| pi-mesons | 57% | 43% |
| antiprotons | 5% | 95% |

$w = 2.7$, accuracy 73%:

| true \ predicted | pi-mesons | antiprotons |
|------------------|-----------|-------------|
| pi-mesons | 51% | 49% |
| antiprotons | 3% | 97% |

Tuning SVM

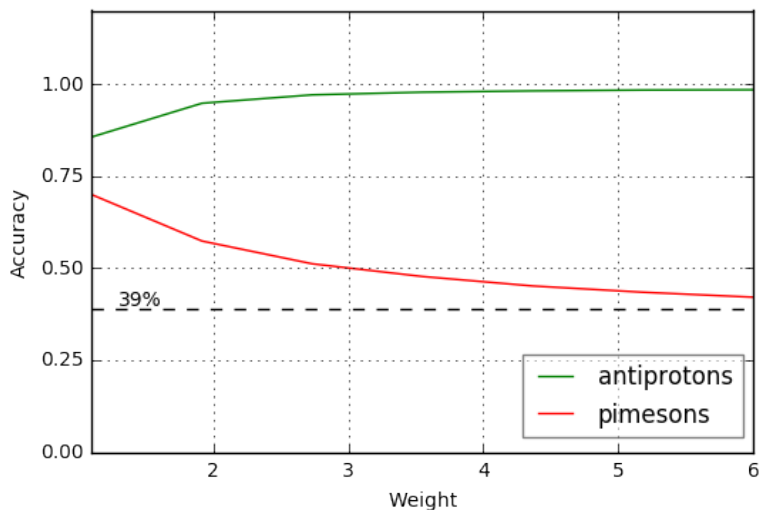
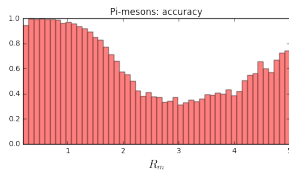
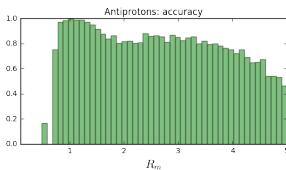


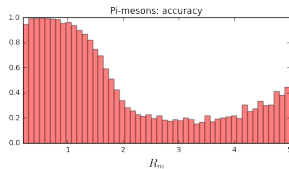
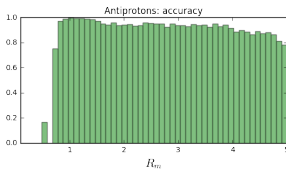
Figure: Classification accuracy of antiprotons and pi-mesons according to increasing class weight for antiprotons.

Unbalanced SVM accuracy

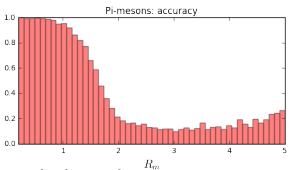
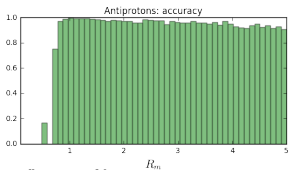
$w = 1.1$, accuracy 78%:



$w = 1.9$, accuracy 75%:

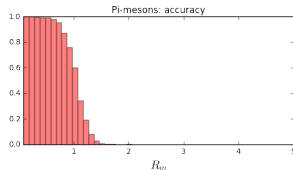
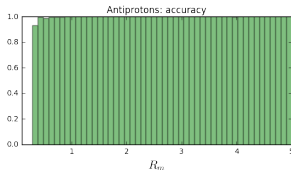


$w = 2.7$, accuracy 73%:

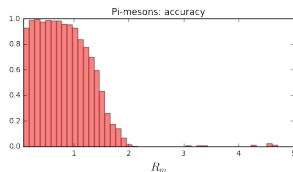
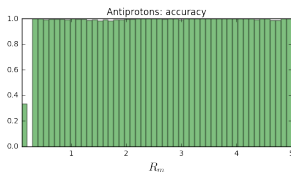


Compared with standard criterion

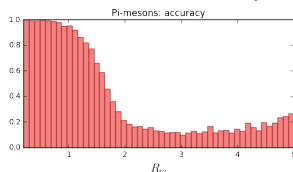
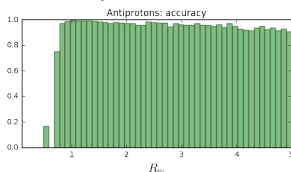
Standard criterion (accuracy 71%):



SVM (track system, accuracy 77%):



SVM (track system + calorimeter, accuracy 73%):



Conclusion

- ▶ SVM classifier shows better results (pi-mesons 51.8%, antiprotons 99.5%, average accuracy 77%) than standard criterion (pi-mesons 39%, antiprotons 99.9%, average 71%).
- ▶ SVM trained on track system features shows classification accuracy of 51% for pi-mesons and 97% for protons, average is 73%.
- ▶ Using only track system features SVM we can classify pi-mesons for R_m up to 2 GV.
- ▶ However, events with $R_m > 2$ GV can't be separated using only track system features.
- ▶ Addition of calorimeter features to descriptor vector allows to classify pi-mesons for $R_m \in [2, 5]$ at approximately 20% accuracy.

Thank you for your attention!