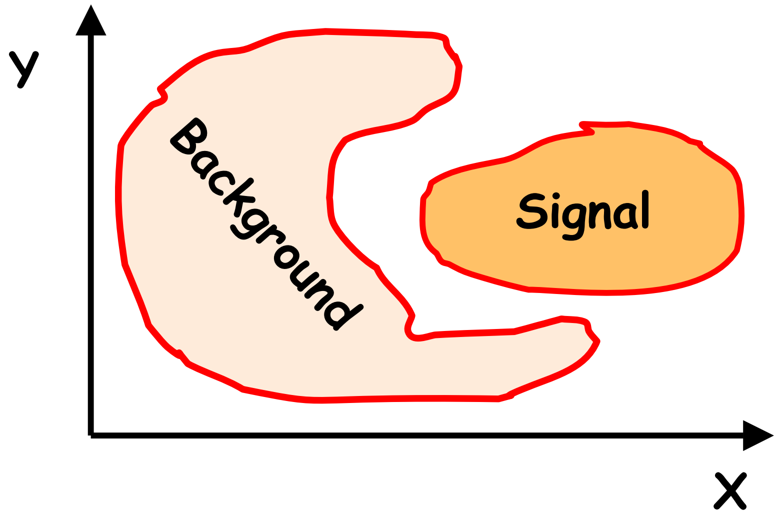
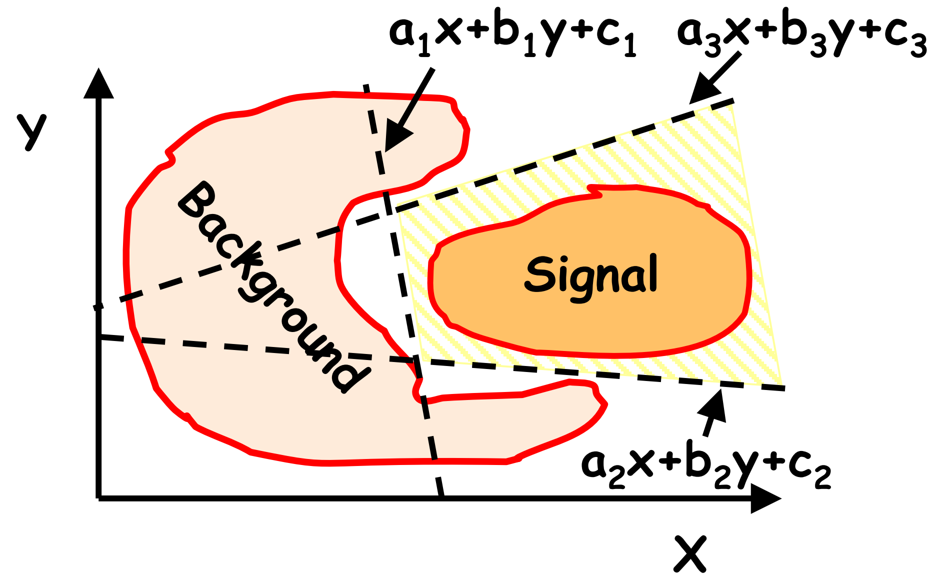
-Methods: Artificial Neural Networks-

- ANN can be trained by MC generated events
- A trained ANN provides multidimensional cuts for data that are difficult to deduce in the usual manner from 1-d or 2-d histogram plots.
- ANN has been used in HEP
- HEP Packages:
  - JETNET
  - SNNS
  - **MLP fit**

-ANN BASICS-



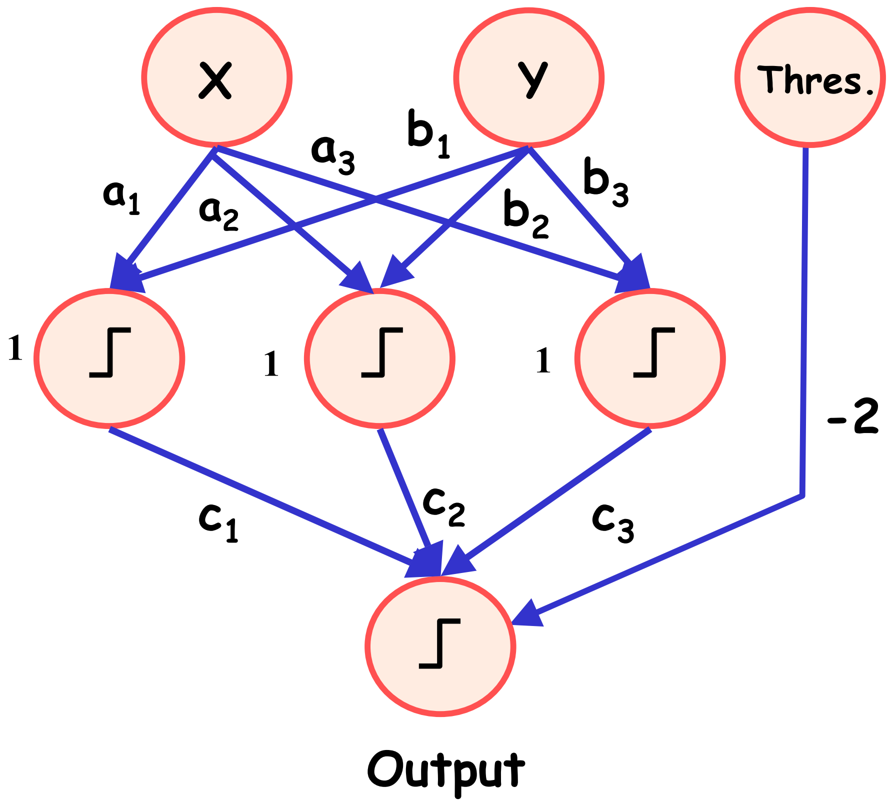


- Event sample characterized by two variables X and Y (left figure)
- A linear combination of cuts can separate "signal" from "background" (right fig.)
- Define "step function" $S(ax+by+c)=\begin{cases}0 & \text{"Signal (x, y)" OUT} \\ 1 & \text{"Signal (x, y)" IN}\end{cases}$
- Separate "signal" from "background" with the following function:

$$C(x,y)=S(S(a_1x+b_1y+c_1)+S(a_2x+b_2y+c_2)+S(a_3x+b_3y+c_3)-2)$$

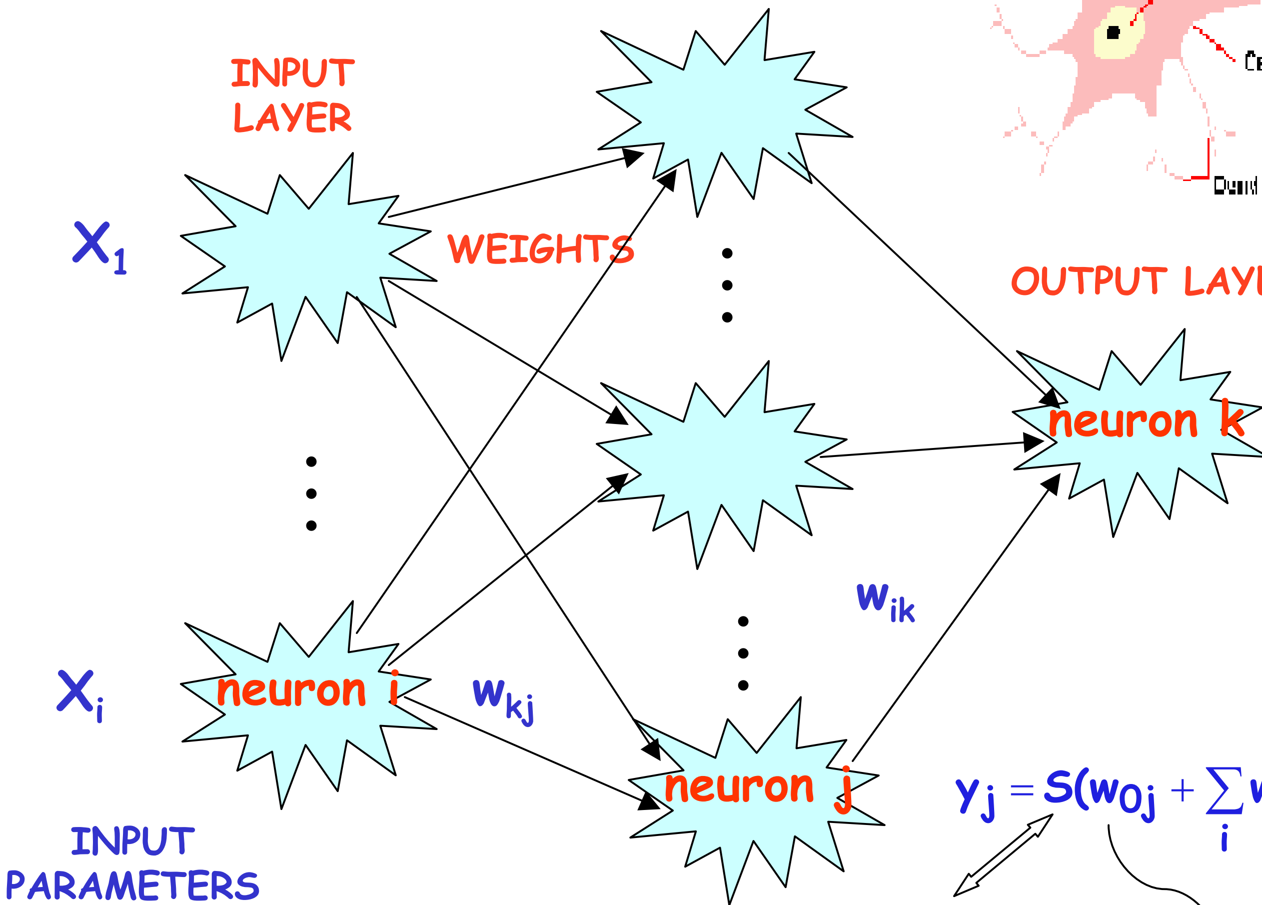
-ANN BASICS-

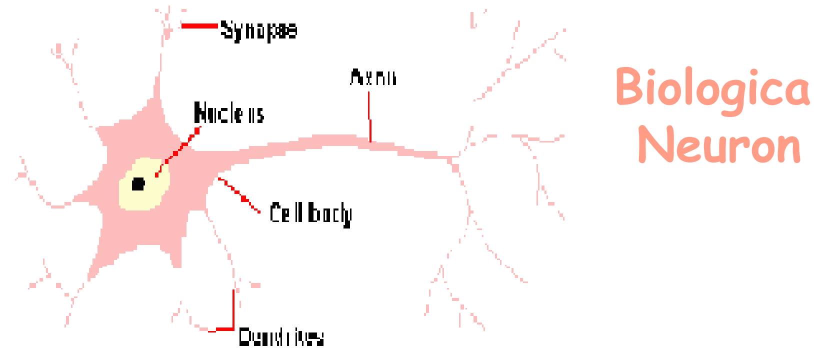
Visualization of function $C(x,y)$



- The diagram resembles a **feed forward neural network** with **two input neurons, three neurons** in the first **hidden layer** and **one output neuron**.
- **Threshold** produces the desired **offset**.
- Constants **a_i, b_i** are the **weights $w_{i,j}$** (i and j are the neuron indices).

-ANN basics : Schematic-





$$t_k = w_{0k} + \sum_j w_{jk} \cdot y_j$$

Bayesian Probability

$$y_j = S(w_{0j} + \sum_i w_{ij} \cdot t_i)$$

Bias

$$S(x) = \frac{1}{1 + e^{-x}}$$

-ANN BASICS-

- **Output** of t_j each neuron in the first hidden layer :

$$t_j = S\left(\sum_i w_{ij} \cdot t_i\right)$$

- **Transfer function** is the sigmoid function :

$$S(x) = \frac{1}{1 + e^{-x}}$$

  - **For the standard backpropagation training procedure of neural networks, the derivative of the neuron transfer functions must exist in order to be able to minimize the network error (cost) function E.**
  - ***Theorem 1 : Any continuous function of any number of variables on a compact set can be approximated to any accuracy by a linear combination of sigmoids***
  - ***Theorem 2 : Trained with desired output 1 for signal and 0 for background the neural network function (output function t_j) approximates the Bayesian Probability of an event being a signal.***

-ANN BASICS-

- Error function : $E = \sum_{p} E_p = \sum_{jp} (d_{pj} - t_{pj})^2$, where
  - p : runs over the events of the training set,
  - j : the index of an output neuron,
  - d_{pj} : the desired output of neuron j in event p,
  - t_{pj} : the network output.

- All **minimization** methods use the computation of first order derivatives:

$$\frac{\partial E}{\partial w_{ji}} = \sum_{p} \frac{\partial E_p}{\partial w_{ji}}$$

- The description of **backpropagation** is that in each iteration :

$$\Delta_p w_{ji}(n+1) = -\varepsilon \frac{\partial E_p}{\partial w_{ji}} + a\Delta_p w_{ji}(n)$$, where

  - $\Delta_p w_{ji}(n+1)$: the **change in w_{ji}** in iteration n+1,
  - ε : the distance to move along the gradient ('**learning coefficient**')
  - α : a smoothing term ("**momentum** ")

-ANN Probability (review)-

ANN analysis : Minimization of an Error (Cost) Function

$$E_N = \frac{1}{N}\sum_N (f(x_i,w) - t_i)^2, \; w = \text{weights}, f(x,w) = \text{ANN output}, x = \text{feature vector}$$

$$t = \text{desired ANN output (1 Signal \& 0 background)}$$

$$E_N = \frac{N_S}{N}\frac{1}{N_S}\sum_S (f-1)^2 + \frac{N_B}{N}\frac{1}{N_B}\sum_B (f-0)^2$$

$$\lim_{N,N_S,N_B\to\infty} E_N = \lim_{N,N_S,N_B\to\infty} \left(\frac{N_S}{N}\frac{1}{N_S}\sum_S (f-1)^2 + \frac{N_B}{N}\frac{1}{N_B}\sum_B (f-0)^2\right)$$

$$\text{but} \lim_{N,N_S\to\infty}\frac{N_S}{N} = P(S) \; \& \lim_{N,N_B\to\infty}\frac{N_B}{N} = P(B)$$

$$\text{and} \lim_{N_S\to\infty}\frac{1}{N_S}(f-s)^2 = \int (f-s)^2 P(x/S)dx \; .....$$

$$...... f = P(S/x)$$

The ANN output is the Bayes a posteriori probability & in the proof no special assumption has been made on the a priori P(S) and P(B) probabilities (absolute normalization)….. TRUE BUT THEIR VALUES DO MATTER ………(They should be what nature gave us)

-ANN probability (review)-

- **Bayesian a posteriori probability :**

$$P(S/x) = \frac{P(x/S) * P(S)}{(P(S) * P(x/S) + P(B) * P(x/B)}$$

$P(S) = \text{a priori} \quad \text{signal probabilty}$ $\qquad P(x/S) = \text{Signal probabilty density function}$

$P(B) = \text{a priori} \quad \text{background probabilty}$ $\qquad P(x/B) = \text{Background probabilty density function}$

- **ANN output : P(S/x)**
- **ANN training examples : P(x/S) & P(x/B)**
- **ANN number of Signal Training Examples P(S)**
- **ANN number of Background Training Examples P(B)**

The MLP (ann) analysis and the Maximum Likelihood Method (Bayes Classifier) are equivalent.

(c_{11} c_{22} = cost for making the correct decision &

c_{12} c_{21} = cost for making the wrong decision)

$$\Lambda(x) = \frac{P(x/S)}{P(x/B)} \; \& \; \xi = \frac{P(B)(c_{12} - c_{11})}{P(S)(c_{21} - c_{22})}$$

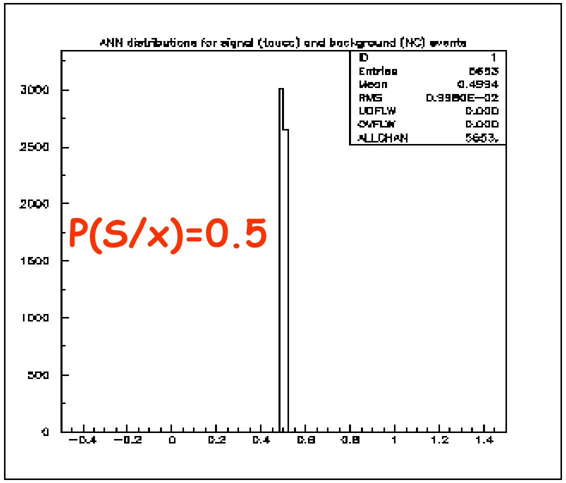
$$\text{if} \quad c_{11} = c_{22} = 0 \; \& \; c_{12} = c_{21} \Rightarrow$$

$$\Lambda(x) > \xi \Leftrightarrow \frac{P(x/S)}{P(x/B)} > \frac{P(B)}{P(S)} \Leftrightarrow P(x/S) * P(S) > P(x/B) * P(B) \Leftrightarrow$$

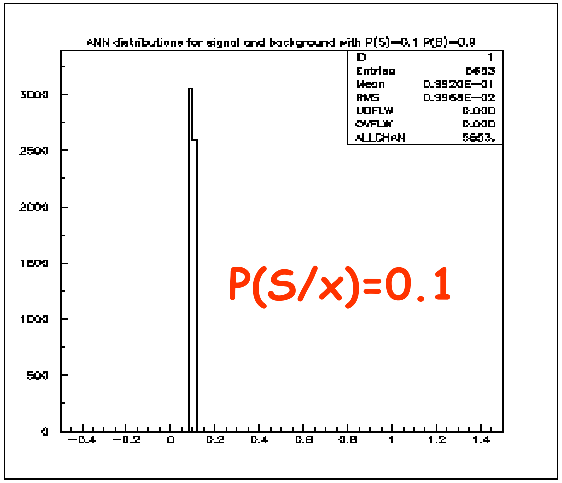
$$\Leftrightarrow \frac{P(x/S) * P(S)}{P(x)} > \frac{P(x/B) * P(B)}{P(x)} \Leftrightarrow P(S/x) > P(B/x) \Leftrightarrow$$

$$\Leftrightarrow P(S/x) > (1 - P(S/x)) \Leftrightarrow P(S/x) > 0.5$$

-ANN Probability cont.-



ANN output



- **Worse hypothetical case 1:**

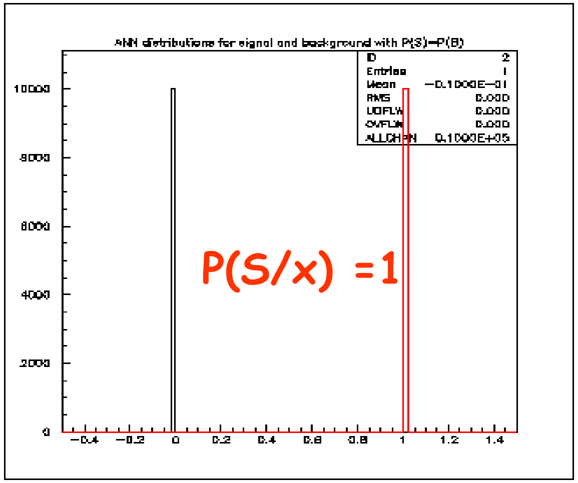
One variable characterizing the populations, which is identical for S and B, therefore :

P(S)=0.1 & P(B)=0.9

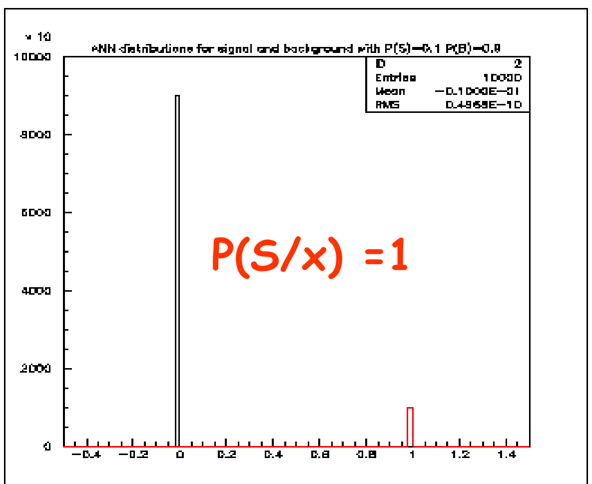
- If we train with equal numbers for signal and background the ANN will wrongly compute **P(S/x)=0.5**

- If we train with the correct ratio for signal and background the ANN will correctly compute **P(S/x)=0.1**, which is exactly what Bayes a posteriori probability would give also.

-ANN Probability cont.-



ANN output



- <u>**Best hypothetical case :**</u>

One variable characterizing the populations, which is completely separated (different) for S and B.

P(S)=0.1 & P(B)=0.9

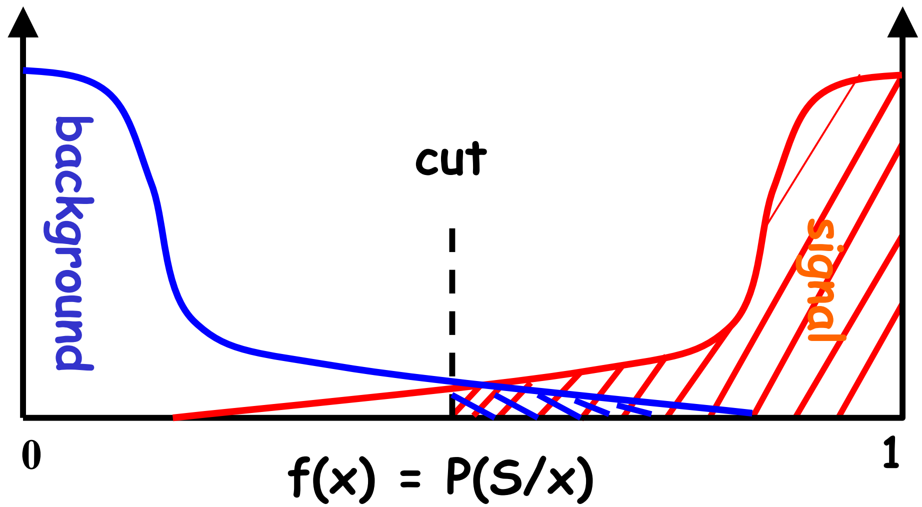
- If we train with equal numbers for signal and background the ANN will compute **P(S/x)=1.**

- If we train with the correct ratio for signal and background the ANN will again compute **P(S/x)=1.**

- **In this case it does not matter if we use the correct a priori probabilities or not.**

-Quantities that characterize an ANN-

Network output (selection) function for “background ”and “signal” events



S = Total # Signal events

B = Total # Background events

S_c = Signal events above Cut

B_c = Background events above Cut

$$\text{efficiency} = \frac{S_C}{S}$$

$$\text{purity} = \frac{S_C}{S_C + B_C}$$

$$\text{contamination} = \frac{B_C}{B}$$

-Goals of the ANN analysis involving spectrometer information -

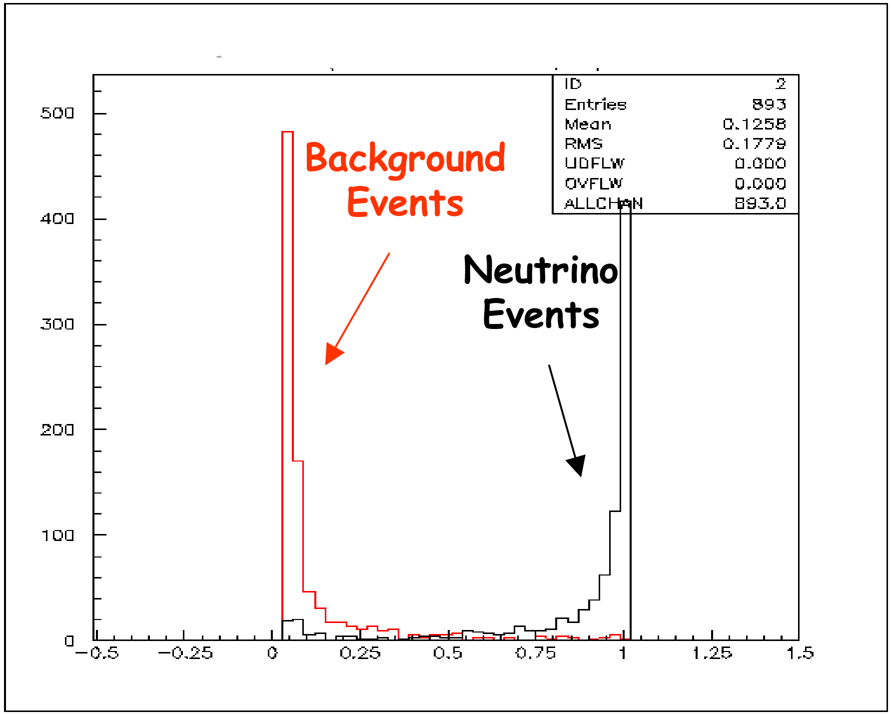
- Use Artificial Neural Network techniques to identify and classify Neutrino Interactions on "event-by-event" basis using topological and physical characteristics of neutrino events derived from both experimental data as well as MC generated interactions:
  - CC ν_μ ν_e ν_τ
  - NC

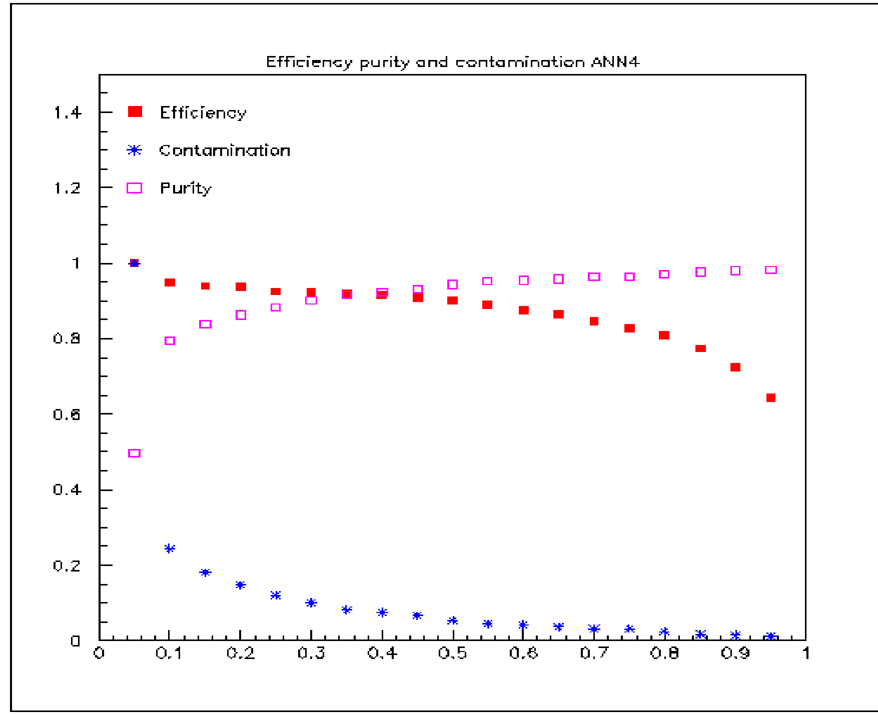
- Requirement: MC should be capable of describing very well the neutrino data.

-ANN Input Variables-

- **Scintillating Fiber System :**
  - Total Number of SF hits (and Total number of "interaction" SF hits 500)
  - Total Pulse height (and Total "interaction" Pulse Height, Pulse height cut @ 500)
  - % of hits in Stations 1 2 3 4 & % of "Interaction hits "
  - Number of SF lines (UZ,VZ)
- **Vector Drift Chambers:**
  - Total Number of VDC hits
- **Drift Chambers:**
  - Total number of DC hits
  - Number of DC tracks
- **EMCAL :**
  - Total Energy Deposition & Total Energy Deposition along $y = 0$ and $| x | > 100$ cm
  - Number of clusters
  - Average cluster energy
  - Mean Cluster angle with respect to the z axis from the interaction point
- **Muon Identification System :**
  - Total number of MID hits
  - Total number of MID hits in the central tubes
- **Other Variables :**
  - Number of 3D final Tracks & Number of 3D final tracks that have SF and DC hits.
  - Trigger Timing Differences (T32,T21,T31)
  - Reconstructed Vertex in the Emulsion Module

-ANN Output Function-

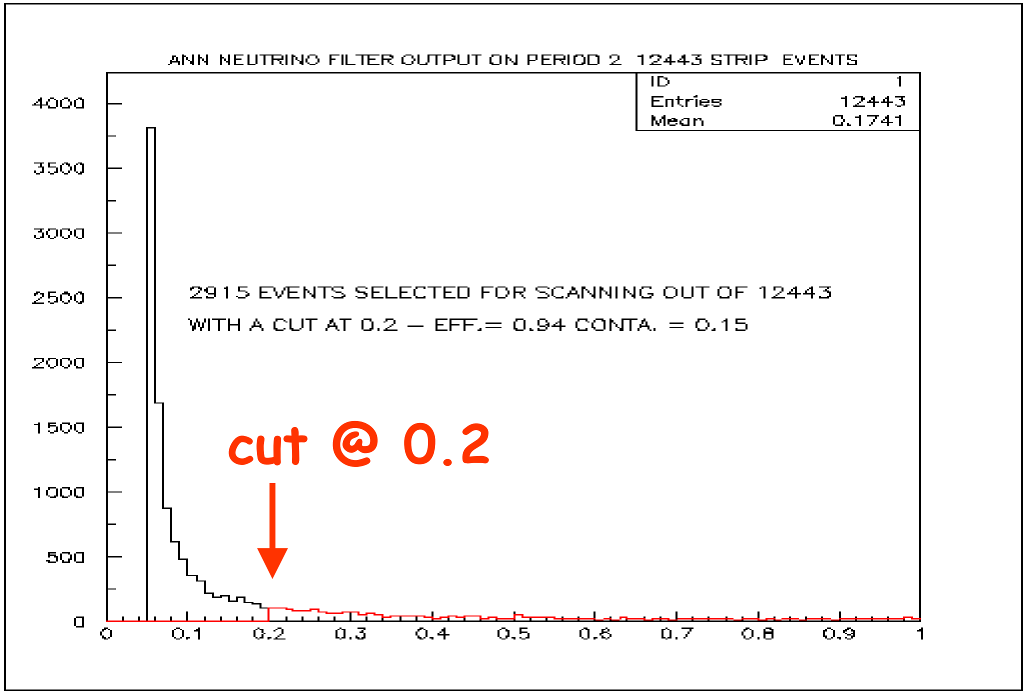




- The **performance** of the ANN is **good** and one can **select events** with **high efficiency** and **high purity (low contamination).**
- With a cut **@ 0.2** :

efficiency 0.94 - purity 0.86 - contamination 0.15

-ANN Implementation & Results on a "raw" Data Sample-



- With a **cut @ 0.2 2915 out of 12443** are **selected** as **"neutrino" interactions.**
- **Initial Signal/Background Ratio ~ 100/12443 = 0.008**
- Obtained Signal/Background Ratio ~ 100/2915 = 0.034

-Neutrino event Classification: Method-

- **Method :**
  - Construct **two sequential** Neural Networks (ANN1 & ANN2) that will be **applied in the whole data set** :

All ν's

a) The first will distinguish ν_μ CC from ν_e & ν_τ CC + NC

b) The second will distinguish NC from ν_e CC & ν_τ CC

-Training Set & Input Variables-

- For **every period** we construct a **separate set of (2) ANN's** since every period has **different target configuration** and thus **different event characteristics**.
- For every period we use **5000 MC** events as a **training set.**

INPUT VARIABLES

HITS Total number of DC hits
(Total number of MID hits in the Central tubes)

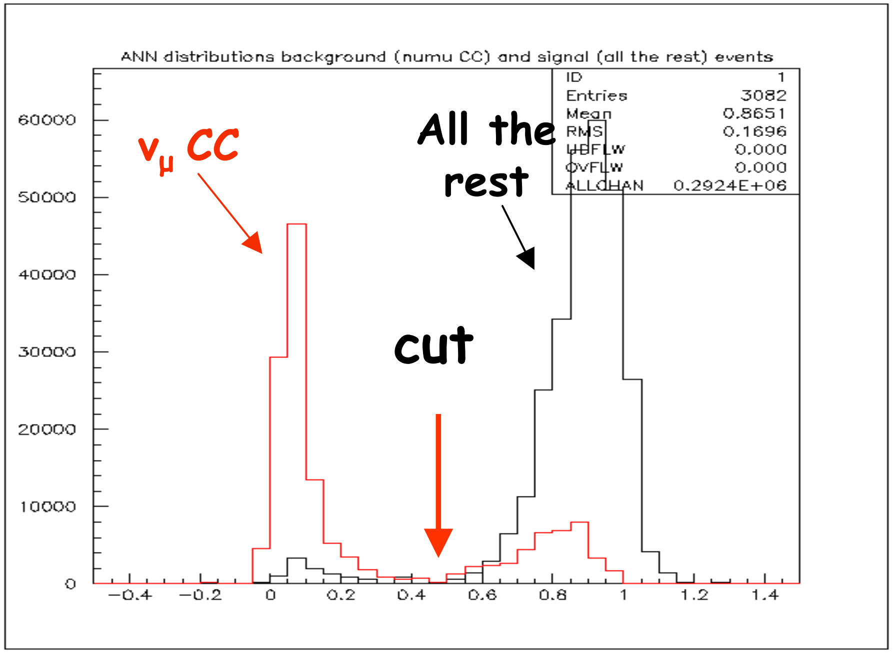
EMCAL Total energy deposition
Number of clusters
Average Cluster energy
Mean value of the Clusters angle from the vertex with respect to the z - axis
Standard deviation of the Clusters angle
Mean Absolute deviation of the of the Clusters angle
Higher Moments of the Clusters angle : a) Skewness b) Curtosis
(Percentage of tracks with E/P < 0.3 (Muons))

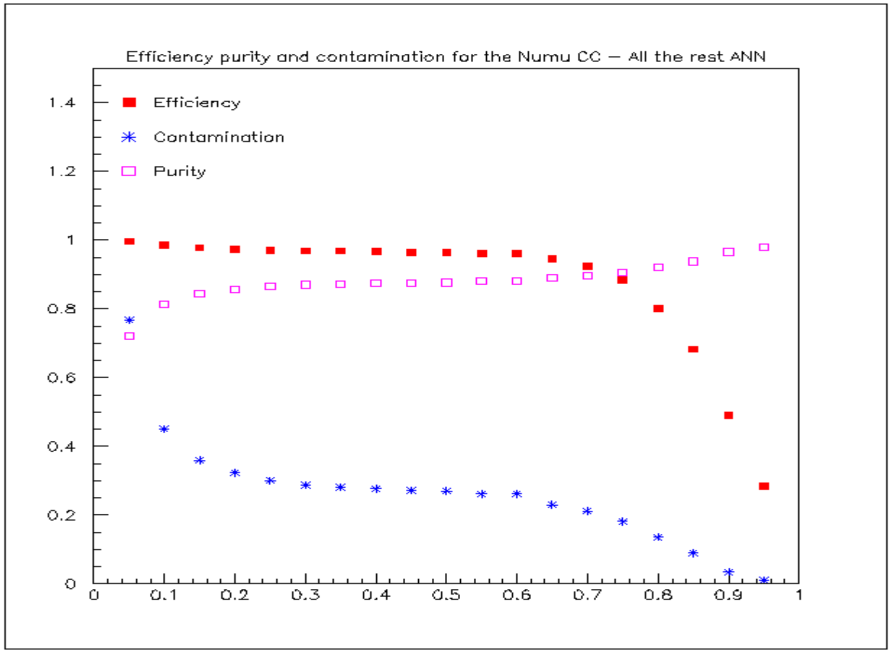
TRACKS Number of final tracks
Number of DC tracks
(Number of tracks that have more than 3 hits in the MID system (Muons))

OTHER Total Pulse Height in the SF system

***** Comparing the MC distributions of these variables with REAL data we found that with the 0.001 criterion they are considered compatible according to the Kolmogorov Test**

-Output of ANN1 (ν_μCC - All the rest)-

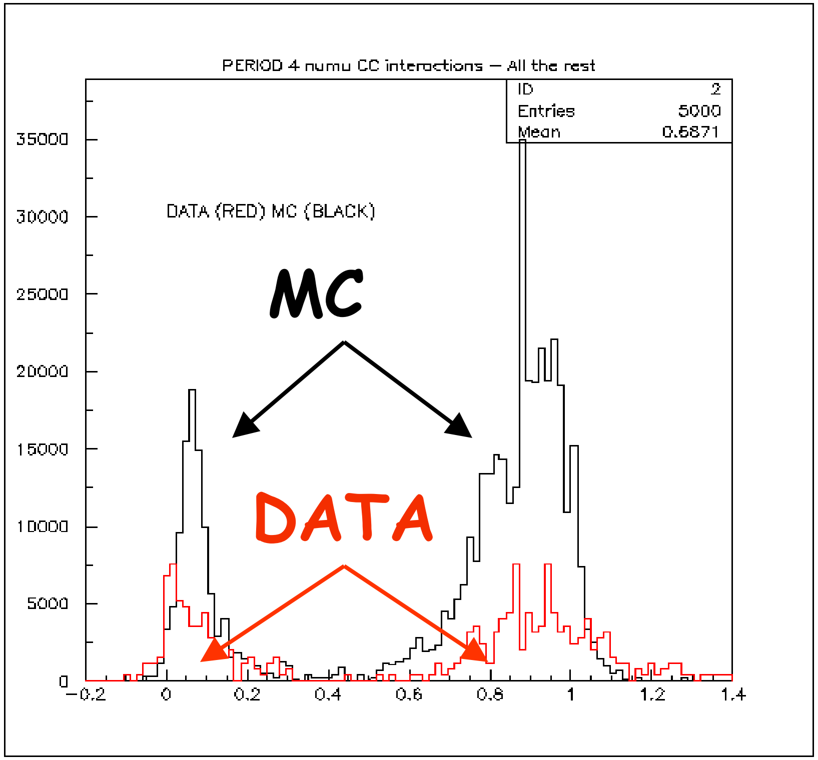




- The **performance** of that network is **satisfactory.**
- With a **cut @ 0.5** in the network output function we select **"signal"** events and on the same time **"background"** events with :

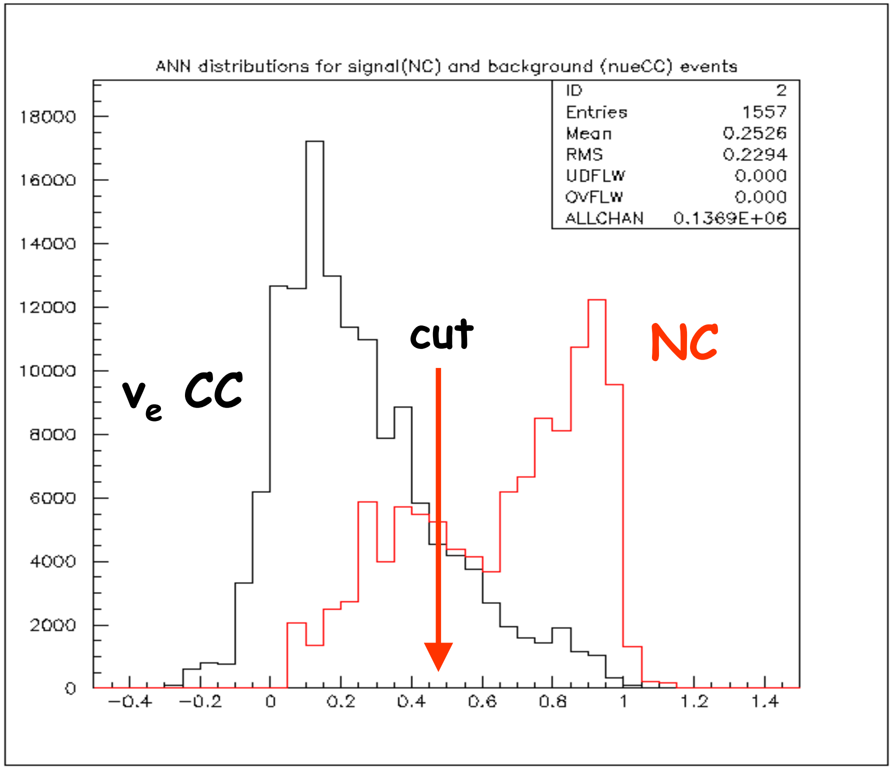
All the rest	**efficiency 96 % - purity 88 %**
ν_μ CC	**efficiency 73 % - purity 96 %**

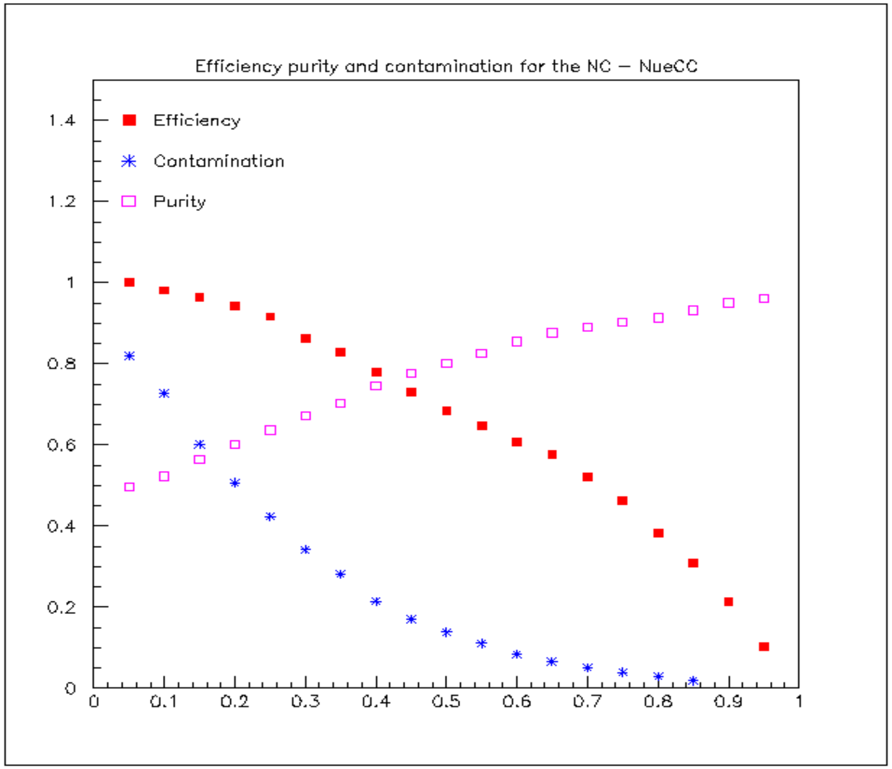
-ANN1 (ν_μCC - All the rest) performance on MC & Real Data-



- The performance of the **output function** of **ANN1** in **MC** events and in the **experimental data set** is **very similar.**
- That indicates that the **results from ANN1** implementation in the **experimental data** set are **quite reliable.**

-Output of ANN2 (NC - ν_e CC)-

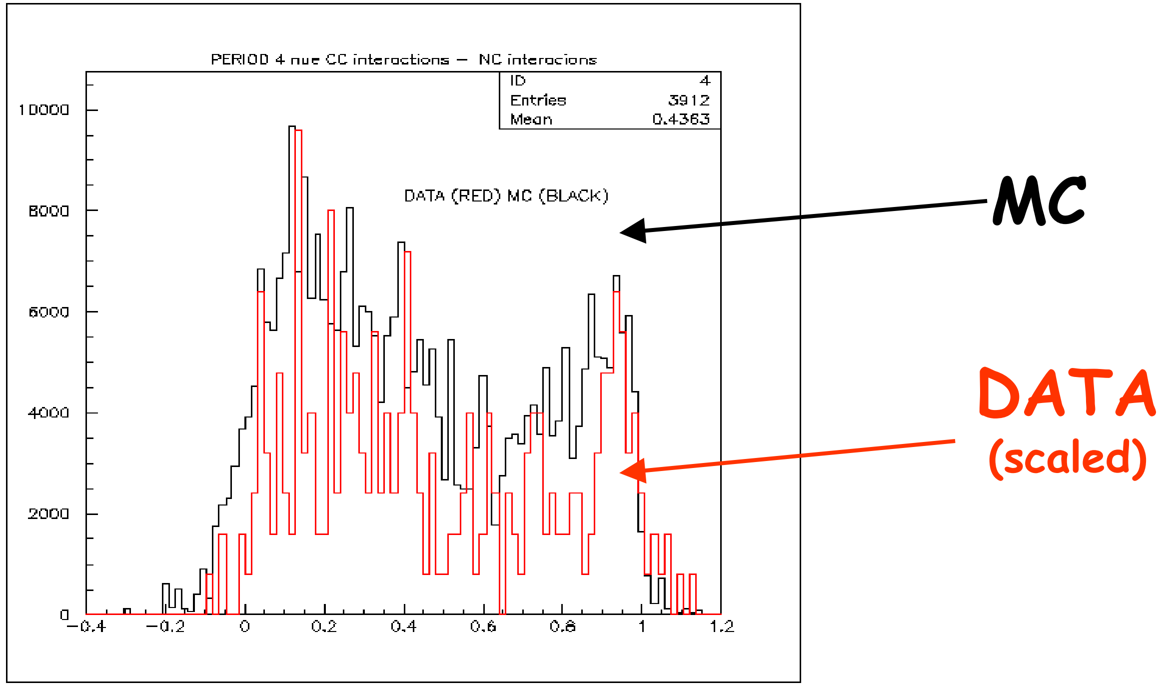




- This network shows a quite good behavior and by choosing a **cut @ 0.5** we select **signal (NC)** and at the same time **background** events **(ν_e CC)** with :

NC	efficiency **68 %** - purity **80 %**
ν_e CC	efficiency **86 %** - purity **76 %**

-ANN2 (NC - ν_e CC) performance on MC & experimental Data-



- The **performance** of the **output function** of **ANN2** in **MC** and in the **Experimental data** set is **very similar**.
- That permits us to consider the **results** of **ANN2** quite **reliable**.

-Expected number of neutrino interactions per run period & per emulsion module-

$$N_{exp.} = \frac{N_{\nu}}{POT} \cdot POT \cdot P_{int.} \cdot \varepsilon$$

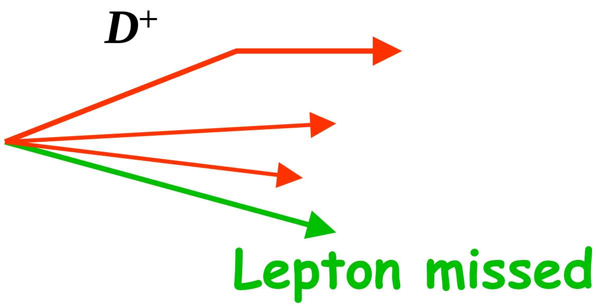
Good agreement (within ~ 1 σ)

Expected number	964 ± 235
Observed number	909
Difference	55 ± 235

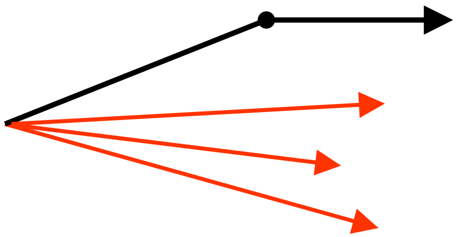
Ratios (%)	ν_{μ} CC	ν_e CC	ν_{τ} CC	NC
Expected	40.9±4.2	32.9±4.0	3.2 ±1.0	22.9±0.1
ANN 'expected'	32.3±2.4	36.3±3.9	------	31.4±2.0
ANN observed	34.3±1.6	36.0±1.6	------	29.7±1.5
Difference	2.0 ±2.9	0.3 ±4.0		1.7 ±2.5
Numbers	ν_{μ} CC	ν_e CC	ν_{τ} CC	NC
Expected	395 ±118	317 ±72	31±11	222 ±37
ANN 'expected'	312 ± 92	350 ±79	------	303 ± 75
ANN observed	312 ± 15	327 ±15	------	270 ± 15
Difference	0 ± 93	23 ± 80		33 ± 76

- Signal & Background -

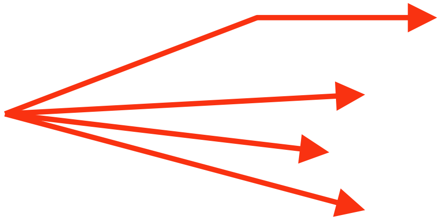
- Charm background



- Interactions (scattering)



- Tau signal

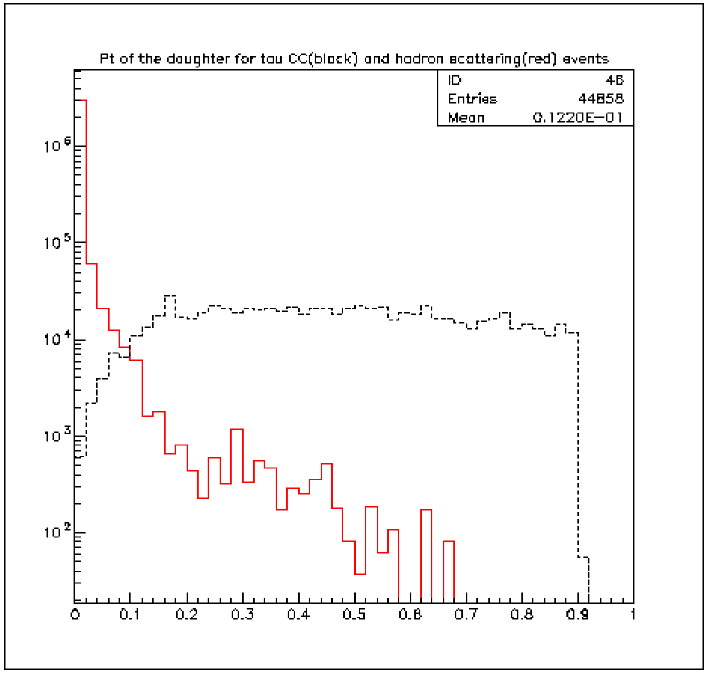


-ANN for ν_τ CC - NC scattering-

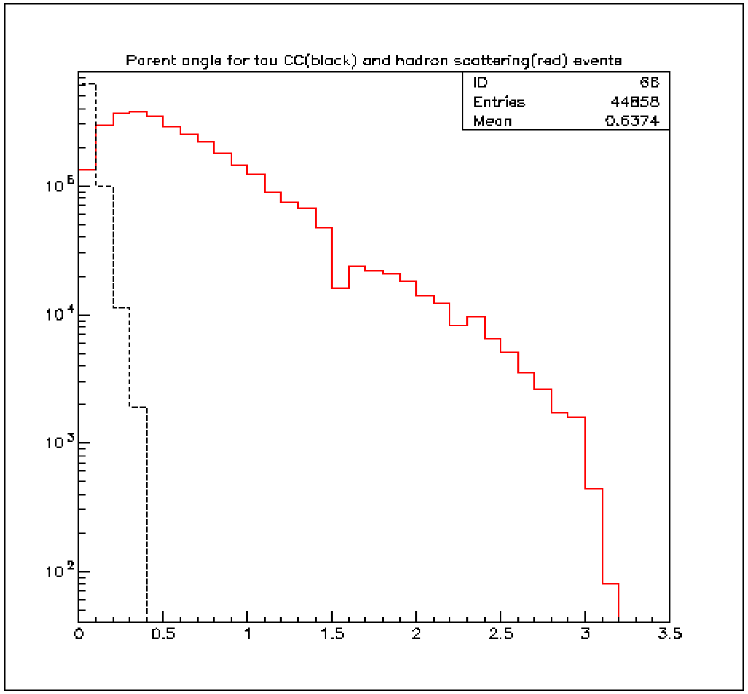
- **Goal :** To separate ν_τ CC interactions from hadron scattering from NC interactions with the use of ANNs
- **Input Variables :**
  - Daughter Momentum
  - Decay Length
  - Parent angle
  - Daughter angle
  - $\Delta\varphi$ (between the parent and all the other primary tracks)
- **Training Set :**
  - 20000 ν_τ CC interactions
  - 20000 hadron scattering NC interactions

-MC Distributions of ν_τ CC & hadron scattering events-

Daughter P_T



Parent Angle



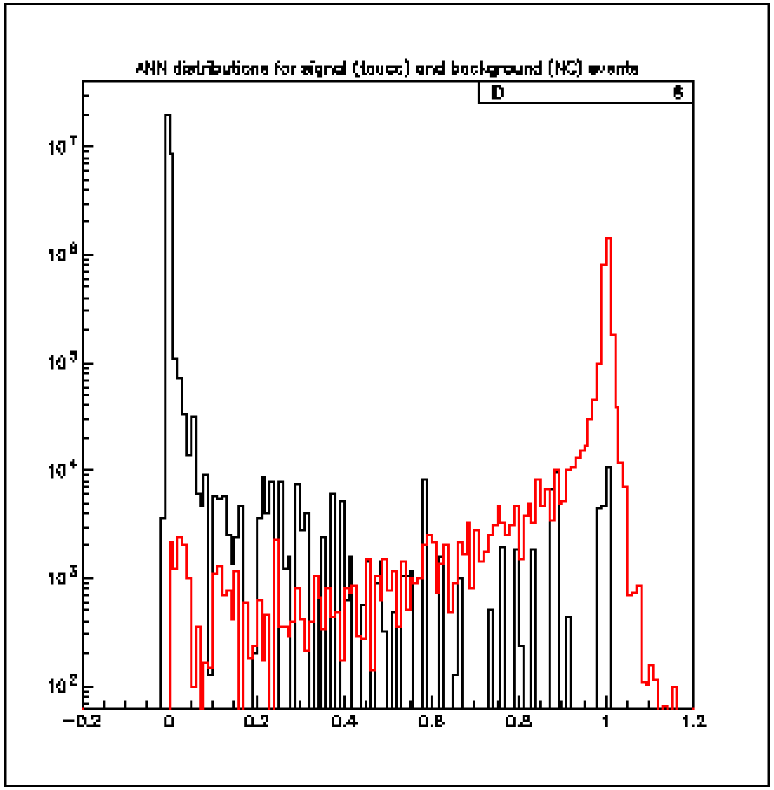
ν_τ *CC* (black)

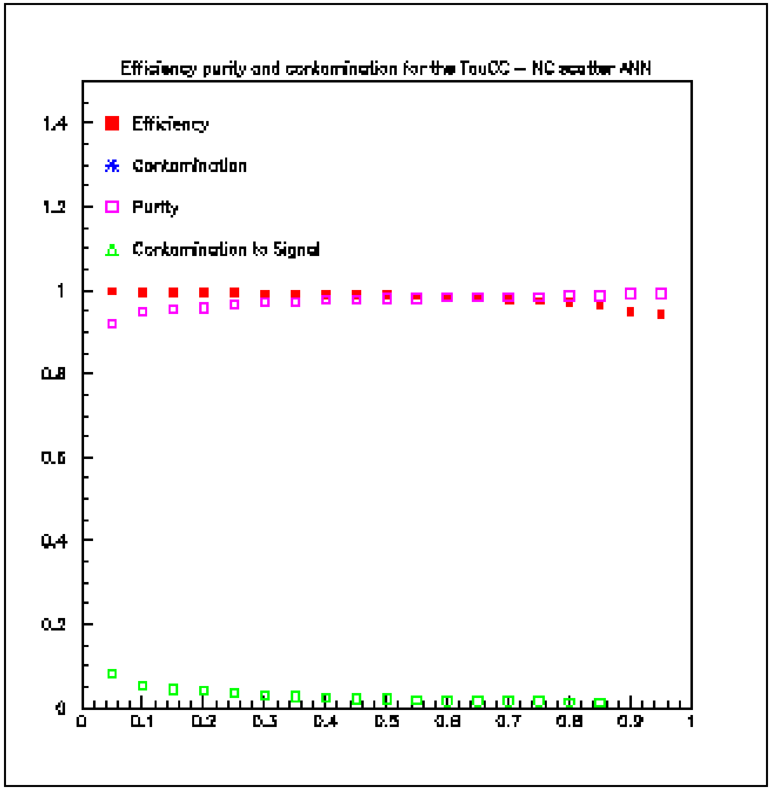
Hadron Scattering (red)

-ANN ν_τ CC - hadron scattering cont.-

Output ANN function (in log scale) (momentum smeared by 30%)

Efficiency, Purity and contamination

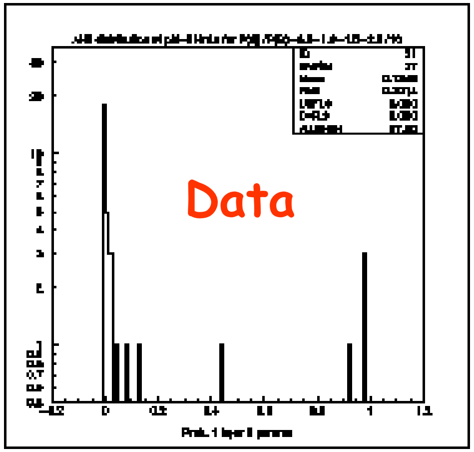




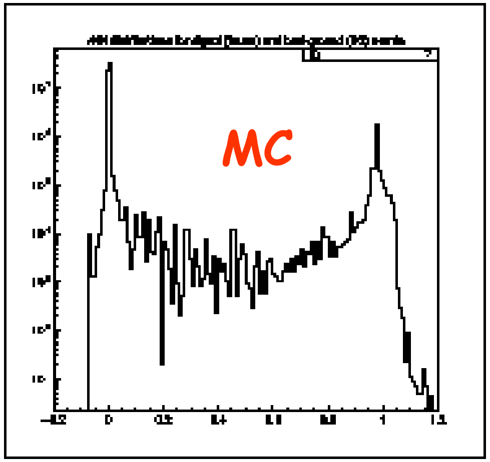
•The performance of the ANN is quite satisfactory as far as its discriminating power is concerned. With the cut@ 0.5 we select tau decays with

~99% efficiency & ~99% purity

-ANN ν_τ CC - hadron scattering results on the 37 recognized kinks-



$\Delta p/p=30\%$



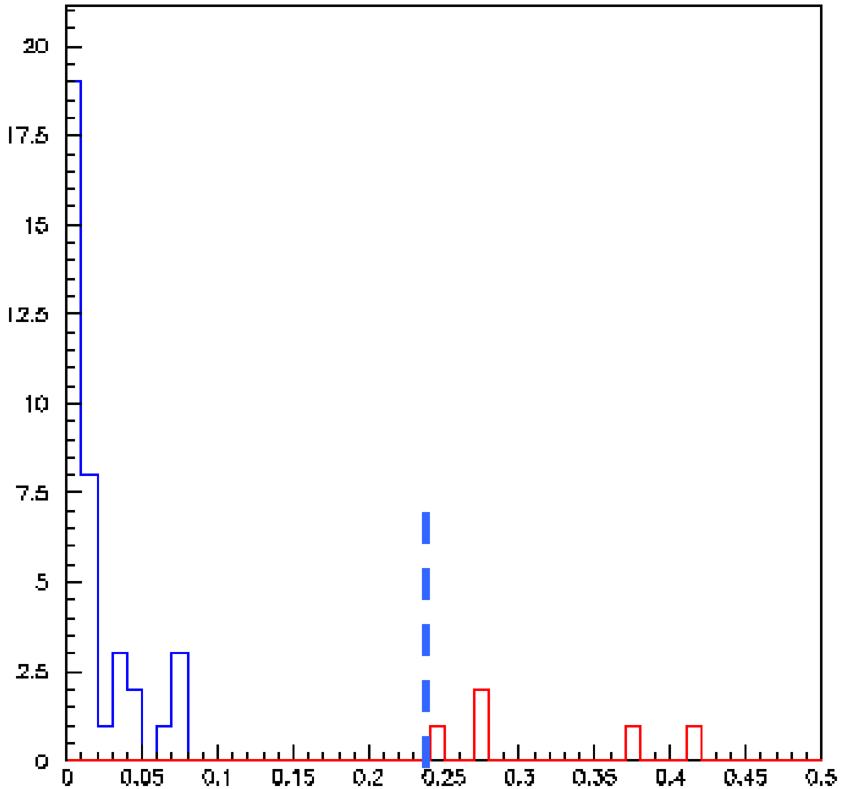
EVENTS THAT EXCEED THE 0.5 CUT IN THE ANN OUTPUT FUNCTION

RUN	EVENT	P_d	θ_d	P_T	L_d	θ_p	$\Delta\varphi$	Probabilities
3263	25102	1.900	0.1300	0.247	1890.1	0.1772	0.176	0.136***
3024	30175	2.900	0.0936	0.271	4504.8	0.0279	1.027	0.971
3039	1910	4.600	0.0895	0.412	276.5	0.0653	2.684	1.000
3333	17665	21.400	0.0130	0.278	564.6	0.0154	2.806	1.000
3193	1361	20.000	0.0187	0.374	1863.6	0.0838	2.341	1.000 CHARM

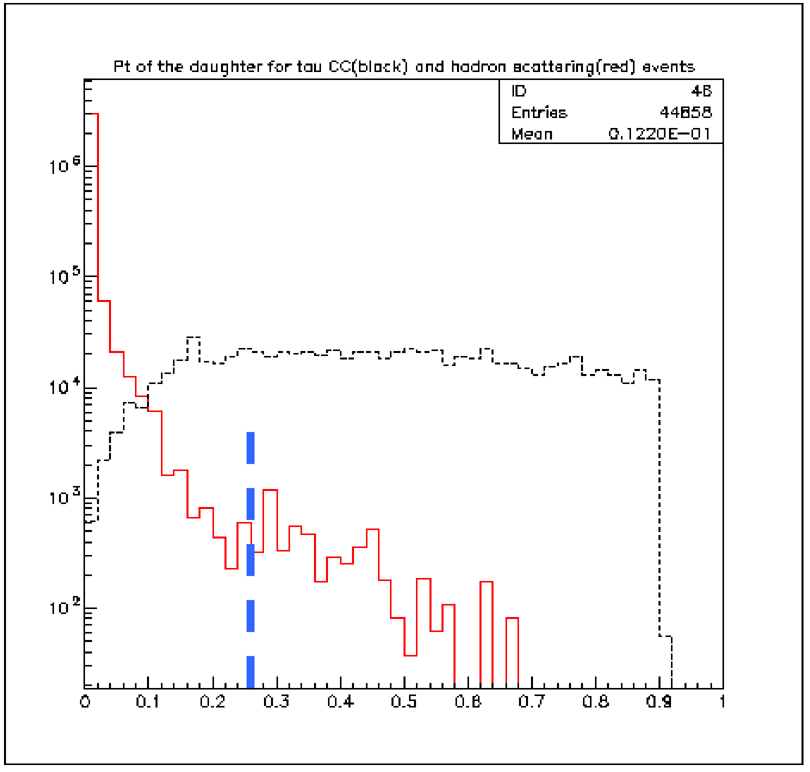
- Considering as "Signal" events (ν_τ CC) the ones with probabilities $P > 0.5$ we can compute the background to these events by adding 1-P. Therefore :

Bkg = 0.029

- Characteristics of Selected ANN events-

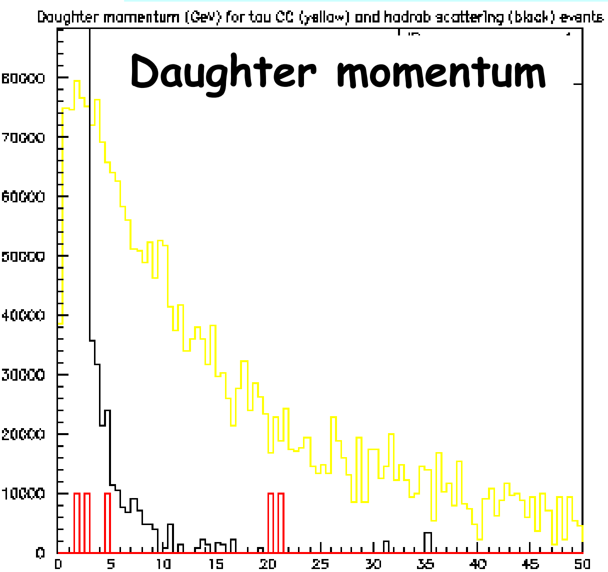


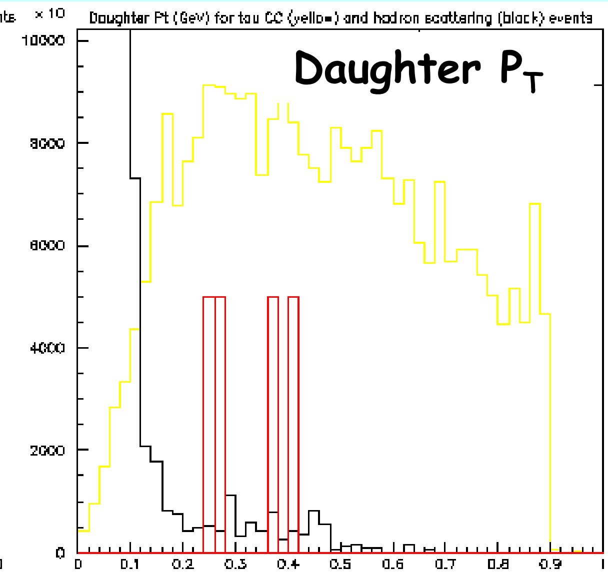
P_T of experimental kinks

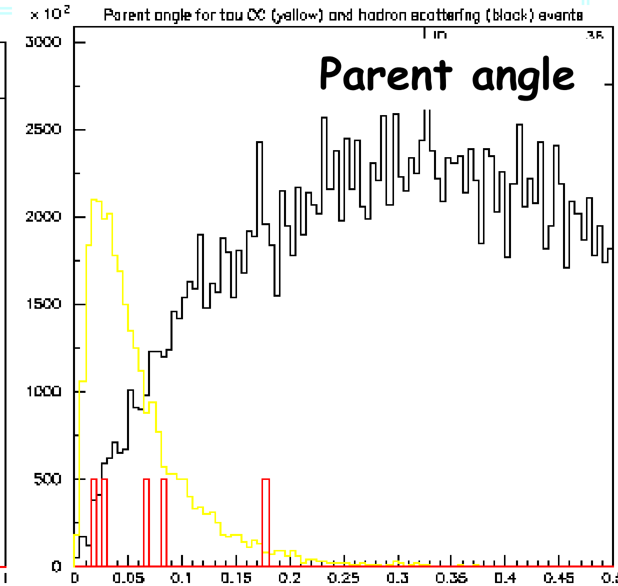


P_T of MC kinks for hadron scattering events (red) and tau decays (black)

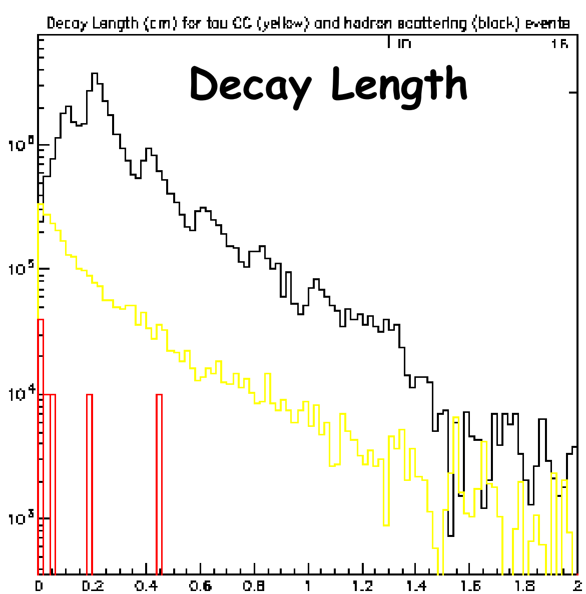
- Characteristics of Selected ANN events cont.-

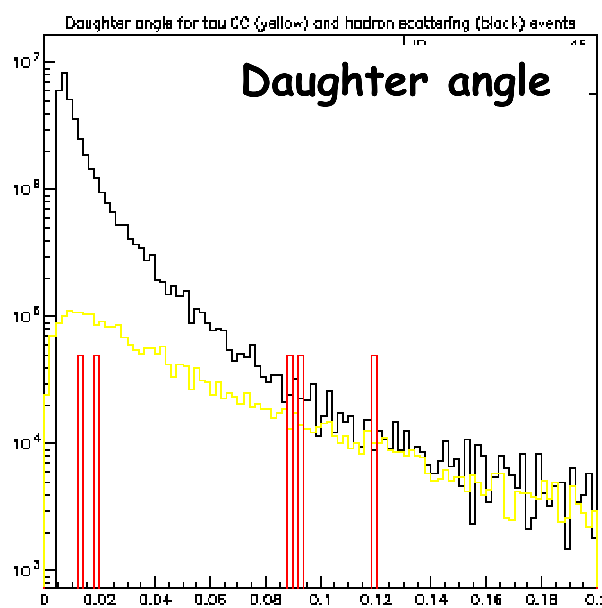


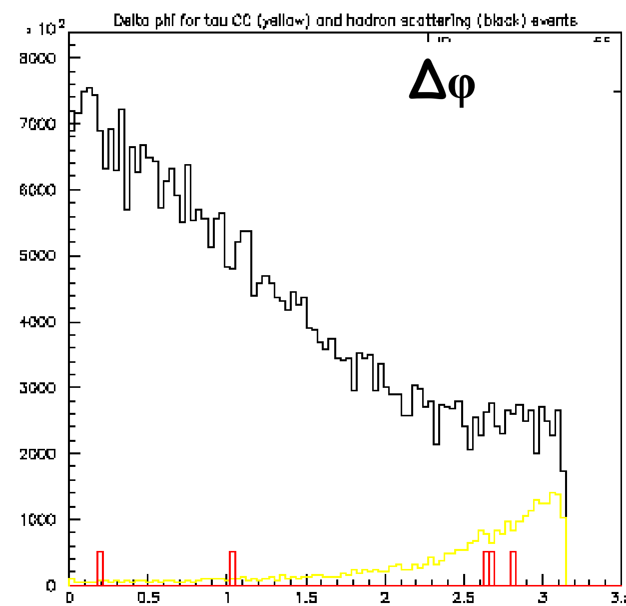




MC, Hadron scattering: Black MC, Tau decays: Yellow Data, Selected candidates : Red

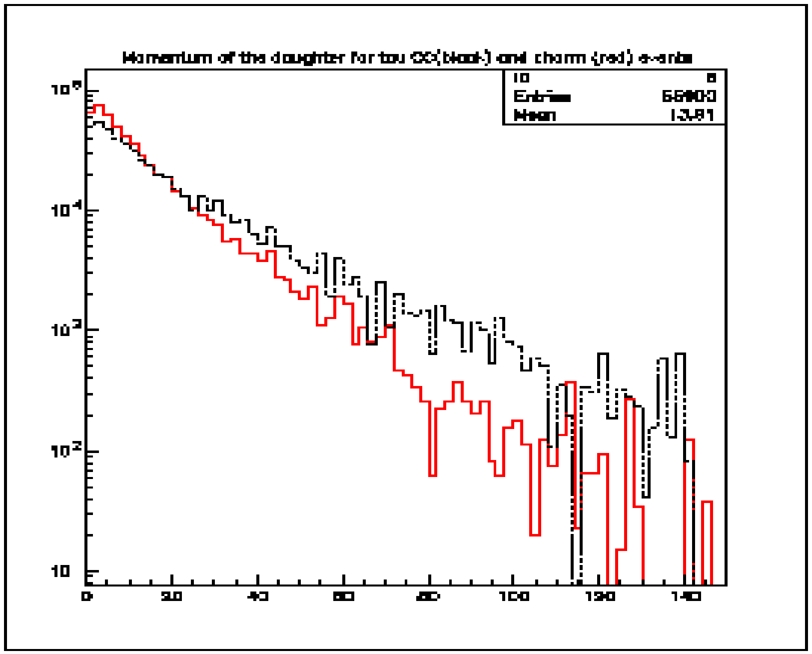




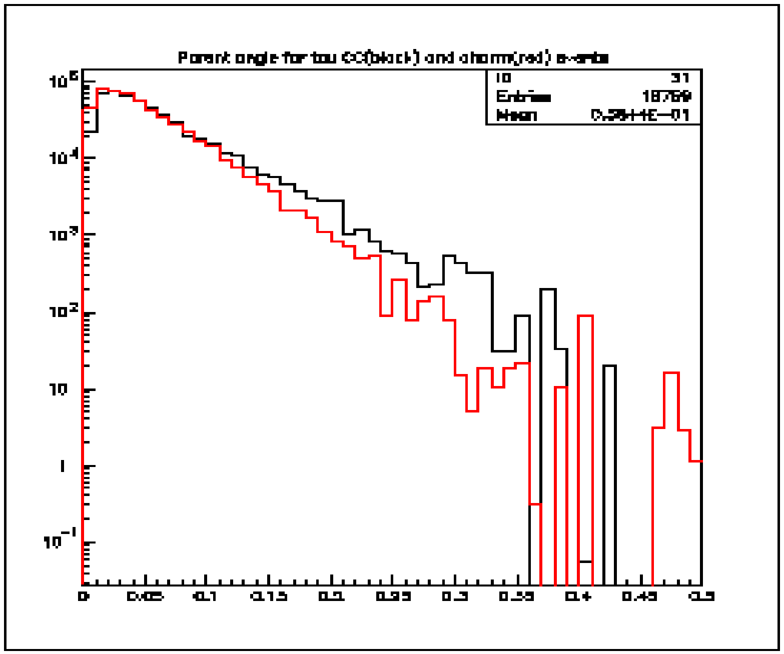


-MC Distributions of ν_τ CC & Charm events-

Daughter Momentum



Parent angle

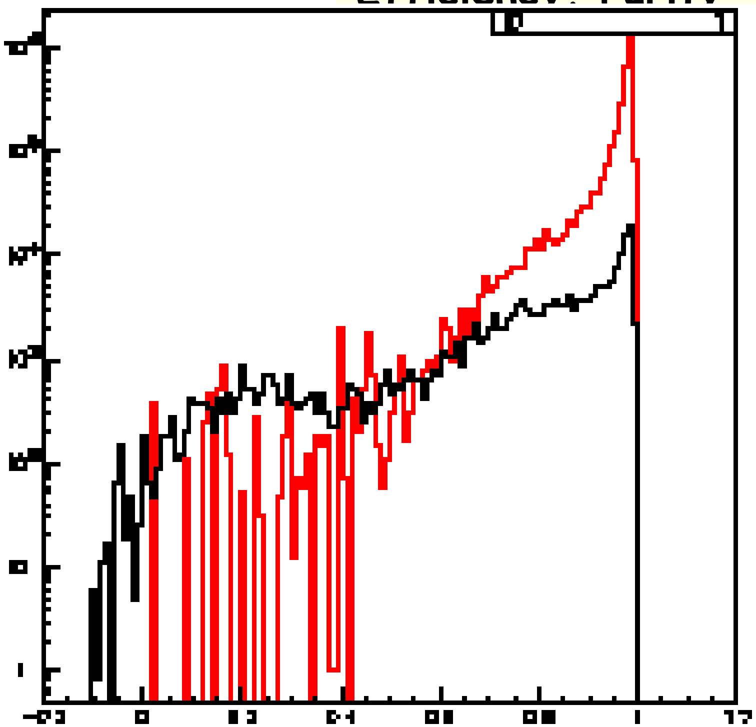


Charm one prong kink decays (red)

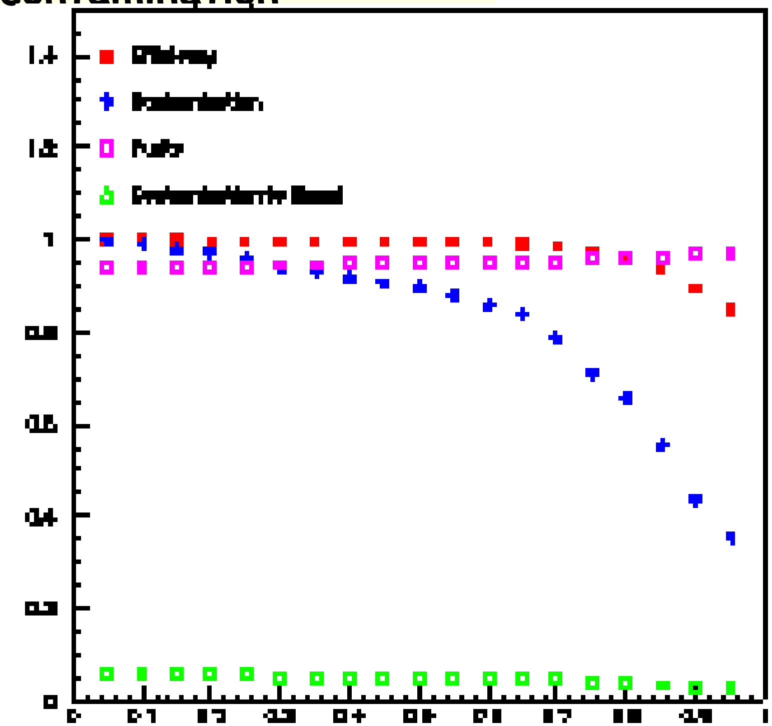
ν_τ *CC* (black)

-ANN ν_τ CC - charm one prong kink decay-

Output ANN function (in log scale) (momentum smeared by 30%)



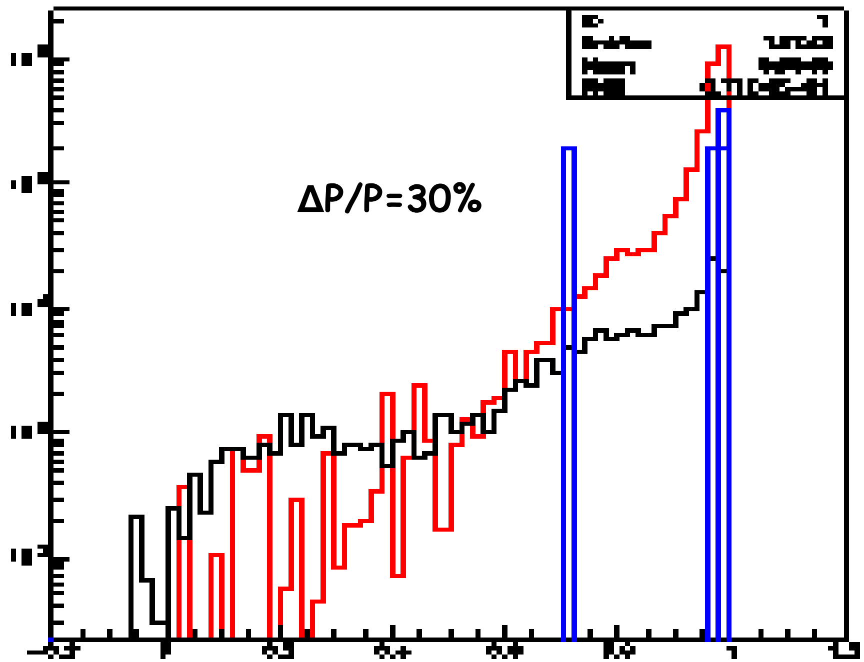
Efficiency, Purity, and contamination



• The classification is poor (as expected), since all variables characterizing these two populations are almost identical.

• However the event probabilities obtained from this ANN analysis can be used to compute the background from this second source (charm one prong kink decays where the lepton from the primary is missed)

-ANN ν_τ CC - charm one prong kink decay background estimation-



RUN	EVENT	P_d	θ_d	P_T	L_d	θ_p	$\Delta\varphi$	Prob.	
3024	30175	2.900	0.0936	0.271	4504.8	0.0279	1.027	0.710	
3039	1910	4.600	0.0895	0.412	276.5	0.0653	2.684	0.990	
3333	17665	21.400	0.0130	0.278	564.6	0.0154	2.806	0.990	
3193	1361	20.000	0.0187	0.374	1863.6	0.0838	2.341	0.990	CHARM

We compute the background to these events by adding 1-P. Therefore :

Bkg = 0.310