Helge Voss: MVA methods and machine learning in HEP GK-Freiburg 5/11/2009

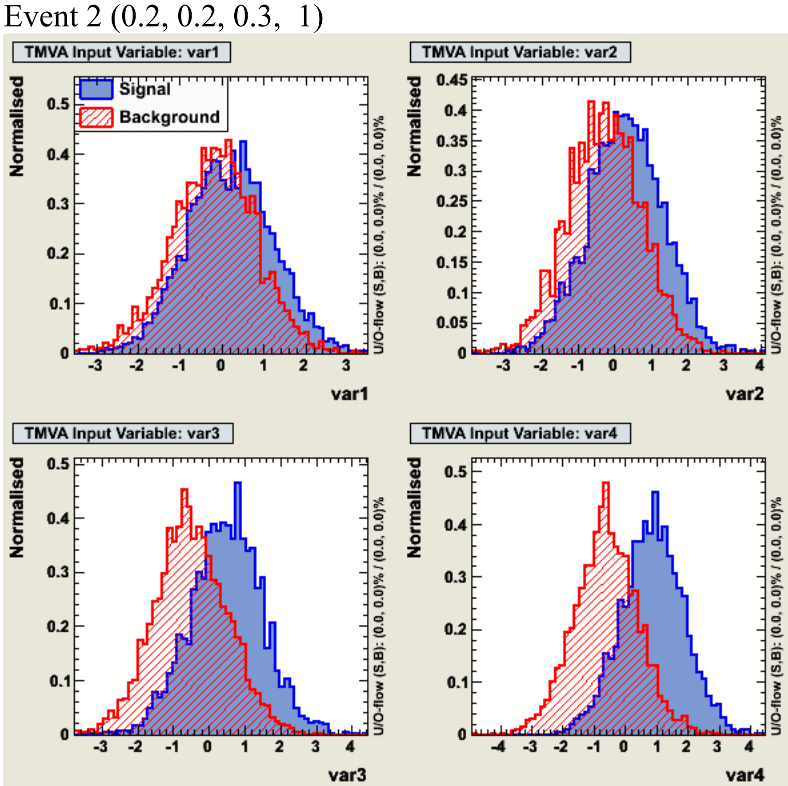
Exercises 1

1. What is the relative error on the number of selected signal events given:

   s = expected number of selected Signal events
   b = expected number of selected Backgr events

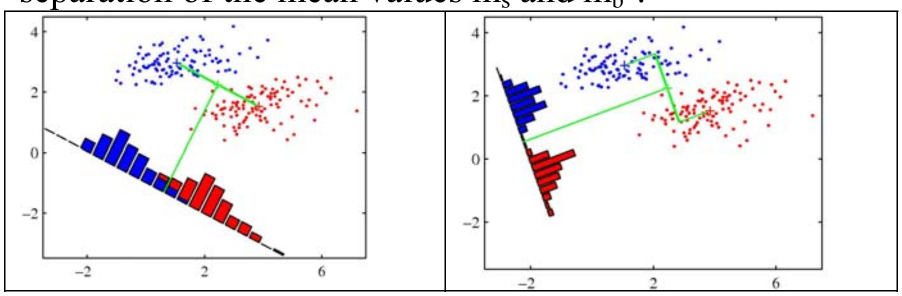
2. Show that the above results is equivalent to maximizing efficiency*purity (or $\sqrt{(\varepsilon * p)}$)
   a. What is the efficiency ε and purity p in terms of N_s, N_{tot}, etc..
   b. Now go on..

3. Which of the two events you'd consider more likely to be signal?
   Event 1 (0.0, -1, 2.5, 2) or
   Event 2 (0.2, 0.2, 0.3, 1)



   b.) Which one of the two is more likely to fail a cut based analysis of signal??
   c.) Looking at b) hat would you propose as possible alternative discriminating technique to cuts?

4. After this warm up: Let's derive Fisher's Linear Discriminant:
   a. Write down the formula for a general linear discriminant y(x)
   b. Sketch a possible shape of the distribution of y(x) for signal events and background events. Or.. better how would you like it to look like for "best" performance ?
   c. What about the mean values of these PDF_S and PDF_B ?
      i. Write down their formula
      ii. How do you maximize their distance ?
      iii. What about $\mathbf{w} \rightarrow a \cdot \mathbf{w}$
      iv. Which one of these two possible y(x) do you think would gives a better separation? Which one has a larger separation of the mean values m_s and m_b ?



      v. Obviously, we need to maximize not only the distance between the mean values, but also …..
   d. Write down "Fishers" criterion of what ("dimensionless") quantity should be maximized to get a nice separation in "y".

e. $J(\vec{w}) = \frac{(m_s - m_b)^2}{(\sigma_S^2 + \sigma_B^2)}$ where $\sigma_{S/B}^2$ is the scatter of S and B events around there respective mean values (in y)

f. Write down σ_S^2 and show that this can be written as: **w** V_s **w,** were $V_s = (V_s)_{ij}$ is the expectation value of the covariance matrix,

g. Using the results and ideas from f) re-write J(w) in the form $J(w) = \frac{\vec{w}^T B \vec{w}}{\vec{w}^T M \vec{w}}$ and determine the matrices B and W

h. What is "special about the matrices B and W. I mean where do we get them from ?

i. In order to find the "best discriminante" we need not to maximize J(w). What's the first "step" that needs to be done if we want to calculate it's maximum ??

j. Yes, of course we need to find w with: $\frac{dJ(\vec{w})}{d\vec{w}} = 0$

   i. First step: write out the derivative (gradient) an transform the equation into the "eigenvalue equation"
   $W^{-1}B \bullet \vec{w} = \lambda \vec{w} \quad \text{where } \lambda = \frac{\vec{w}^T B \vec{w}}{\vec{w}^T W \vec{w}}$
   ii. Show that for any arbitrary vector x, Bx points into the direction of μ_S-μ_B
   iii. What does ii) mean for the eigenvalue equation ?
   iv. Write down the solution