Data set : $D = \{(x_1, y_1, w_1), \ldots, (x_n, y_2, w_2)\}$

$x_i \in \mathbb{R}^d$ "d-dimensional feature vector"

$y_i \in \{b, s\}$: label ("signal", "background")

& $w_i \in \mathbb{R}^+$ non negative weight

define $S = \{i : y_i = s\}$ and $\mathcal{B} = \{i, y_i = b\}$, and $n_s = |S|$, $n_b = |\mathcal{B}|$ number of signal / background events

For the simulated data set:

$n_s / n_b \neq P(y=s)/P(y=b)$ for naturally collected data, as we can simulate as many events as we wish

- this is good because normally $P(y=s) \ll P(y=b)$ → so we would have very few $(y=s)$ points to train our model / i.e. unbalanced data set

importance - weighted events: → to make the setup invariant to the number of events

$$\sum_{i \in S} w_i = N_s \quad \& \sum_{i \in \mathcal{B}} w_i = N_b,$$

where N_s and N_b are normalization constants ≈ "expected total number of s/b events"

$$w_i \sim \begin{cases} p_s(x_i)/q_s(x_i) & \text{if } y_i = s \\ p_b(x_i)/q_b(x_i) & \text{if } y_i = b \end{cases}$$

where $p_s(x_i) = p(x_i \mid y=s)$ & $p_b(x_i) = p(x_i \mid y=b)$ (p: probability density)

q_s & q_b are instrumental densities

define "g" as classifier, "G" $= \{x : g(x) = s\}$ selection region, $\hat{G}$ index set points that g selects

$$\hat{G} = \{i : x_i \in G\} = \{i : g(x_i) = s\}$$

$$s = \sum_{i \in S \cap \hat{G}} w_i$$: unbiased estimator of the expected number of signal events selected by g, $S \cap \hat{G}$: the index set of events captured by the classifier g

$$\mu_s = N_s \int_G p_s(x)\,dx$$ and

$$b = \sum_{i \in \mathcal{B} \cap \hat{G}} w_i$$: unbiased estimator of the expected number of background events selected by g

$$\mu_b = N_B \int_G p_b(x)\,dx$$

in machine learning terminology:

s : true positive

b : false positive

AMS (approximate median significance)

$$AMS = \sqrt{2\left((s+b+b_{reg})\ln\left(1+\frac{s}{b+b_{reg}}\right)-s\right)}$$

where $b_{reg} = cnst = 10$, regularization term to reduce the variance of the AMS

TASK:
- train classifier g to MAXIMIZE AMS!

The data

events of Higgs bosons + 3 background processes

1) Z bosons - similar event to Higgs bosons
2) top quarks
3) W boson decay

importance-sampling flavor:
- the sum of weights of events falling in the region is an unbiased estimate of the expected number of events
- weights are not part of the input to the classifier

$d=30$ features' details are given in Appendix B