

TCGA and five independent cohorts



TCGA+ICGC cohort (n=905)

Expression of 65 genes (P) Patients survival at the third year (s)

Cross validation

Ten-fold cross validation

1 2 3 4 5 6 7 8 9

10

Training set

Validation set

Random forests classifier

Apply classifier to the validation set

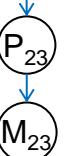
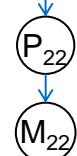
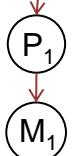
Sorted by importance values

23 genes selected

RALA, PLOD2...

Gene_Expression

...ITPR2, SLC2A2



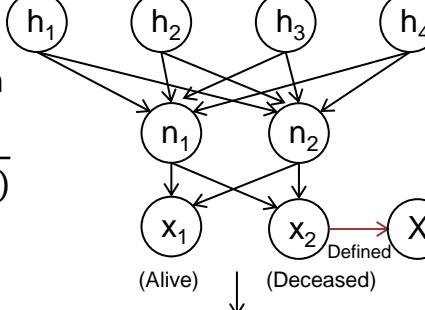
Apply step function and transform to Gene_Score

Softmax function

$$x_k = \frac{\exp(n_k)}{\sum_{i=1}^2 \exp(n_i)}$$

Loss function (F)
(cross entropy error)

$$F = - \sum_{N=1}^{905} t_n \log X_N$$



Minimize loss function by a gradient descent method



Defined and calculate 'Gene_Weight'