Introduction to Computational Medicine I

Project 2: Lecture 1

Objective: to construct predictive models to classify {sepsis, nonsepsis} from EHR and numerical waveforms

Notation: $Y_i \in \{0,1\}$ is label for patient i where sepsis = 1

$$X_i = \begin{pmatrix} x_1^i \\ x_2^i \\ \vdots \\ x_m^i \end{pmatrix}$$ is feature vector for patient i (eg. age, gender)

Problem: Given data set $\{(X_i, Y_i)\}_{i=1}^{N}$ where Y_i are independent samples of a Bernoulli R.V., construct/estimate the PMF

$$\Pr(Y_i = y_0) = (p_i)^{y_0}(1-p_i)^{1-y_0} \qquad y_0 = 0,1$$

where $p_i = f(x_i ; \beta)$

Idea: We want to pick f and β such that $\Pr(Y_1 = y_1, Y_2 = y_2, \ldots, Y_N = y_N)$ is "high". So why not pick f & β to maximize this quantity?

In stastistics, this is called the <u>likelihood function</u>. It is a function of parameters of a statistical model (PMF or PDF) given data. In our case,

$$L(f,\beta) = Pr(Y_1=y_1, Y_2=y_2, \dots, Y_N=y_N; f;\beta)$$

$$= \cancel{\prod_{i=1}^{N} (1-p_i)^{y_i} p_i^{1-y_i}}$$

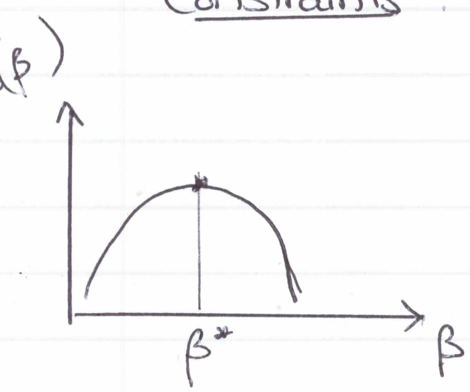
$$= \prod_{i=1}^{N} (p_i)^{y_i} (1-p_i)^{1-y_i}$$

$$= \prod_{i=1}^{N} (f(x_i;\beta))^{y_i} (1-p_i)^{1-y_i}$$

it is not "Probability" r say, it is a nc of a parameter or a given observed outcome!

Q: Why don't we fix $f(\cdot)$-parametrized class of functions?

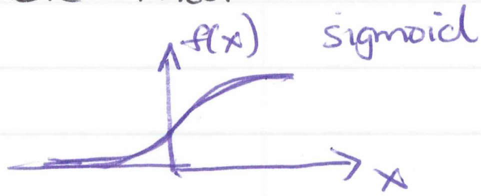
<u>Constraints</u>: (i) need $0 \le f(x) \le 1$
(ii) maximizing L should be doable
(iii) L should be concave in β



Candidate: $$f(x) = \frac{e^{\beta^T x}}{1+e^{\beta^T x}}$$ Satisfies (i)

Does it satisfy (iii)? First note that

$$\log\left(\frac{f(\cdot)}{1-f(\cdot)}\right) = \beta^T x$$



$$\log(f) = \beta^T x - \log(1+e^{\beta^T x})$$
$$\log(1-f) = -\log(1+e^{\beta^T x})$$

$$\Rightarrow \log(f) - \log(1-f) = \beta^T x \quad ✓$$ Back to $L(\cdot)$

$$L(\beta) = \prod_{i=1}^{N} (f)^{y_i} (1-f)^{1-y_i}$$

$$\Rightarrow \log L(\beta) \triangleq \ell(\beta) = \sum_{i=1}^{N} y_i \log(f) + (1-y_i)\log(1-f)$$

$$\Rightarrow \ell(\beta) = \sum_{i=1}^{N} y_i \log\left(\frac{f}{1-f}\right) + \log(1-f)$$

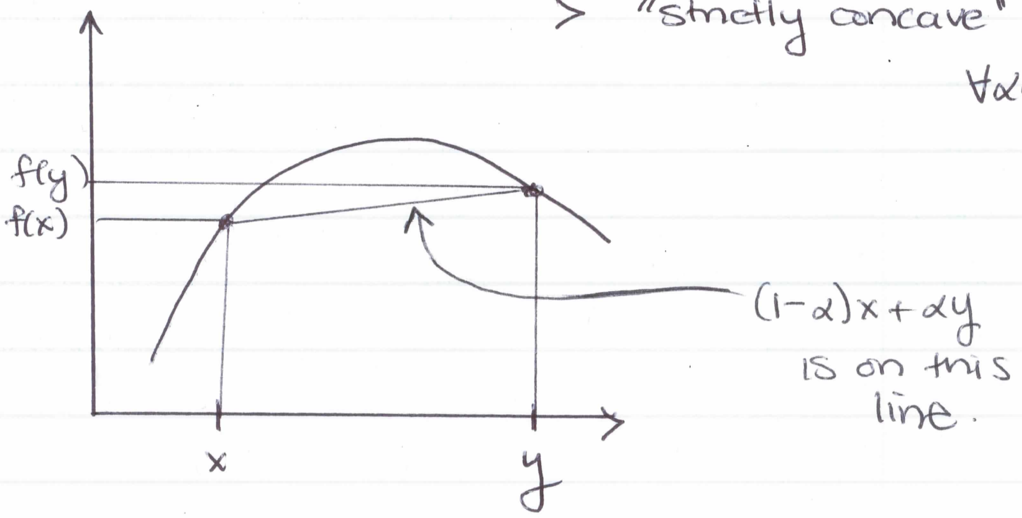
$$\ell(\beta) = \sum_{i=1}^{N} y_i \beta^T x + \log(1-f(\beta))$$

<u>Concave functions:</u> f is concave if

$$f\big((1-\alpha)x + \alpha y\big) \geq (1-\alpha)f(x) + \alpha f(y)$$

$>$ "strictly concave"

$\forall \alpha \in (0,1)$



<u>Fact:</u> if F is differentiable, then

$$f \text{ concave} \overset{\text{iff}}{\iff} f'' \leq 0 \quad \forall x$$

So let's use this fact to check if $\ell(\beta)$ is concave...

$$\ell'(\beta_j) \triangleq \frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{N} y_i x_j - \frac{x_j e^{\beta^T x}}{(1+e^{\beta^T x})} \quad \forall j$$

$$\frac{\partial^2 \ell}{\partial \beta_j^2} = -x_j^2 \frac{e^{\beta^T x}}{(1+e^{\beta^T x})} \leq 0 \quad \forall x_j$$

$\Rightarrow$ $\ell(\beta)$ is concave with our choice of $f(.)$!!

Turns out it is very efficient to compute β^* where

$$\beta^* = \underset{\beta}{\operatorname{argmax}} \; \ell(\beta)$$

Note: We will use "glmfit" in MATLAB.

OK - so now we know how to construct

$$\hat{p}_i = \frac{e^{\beta^{*T} x_i}}{1 + e^{\beta^{*T} x_i}} \qquad i = 1, 2, \ldots, N$$

Q: How do we use $\hat{p}_i$ to classify patient i?

A: if $p_i > \theta$ then decide/classify as sepsis

else classify as nonsepsis

A simple threshold rule!!

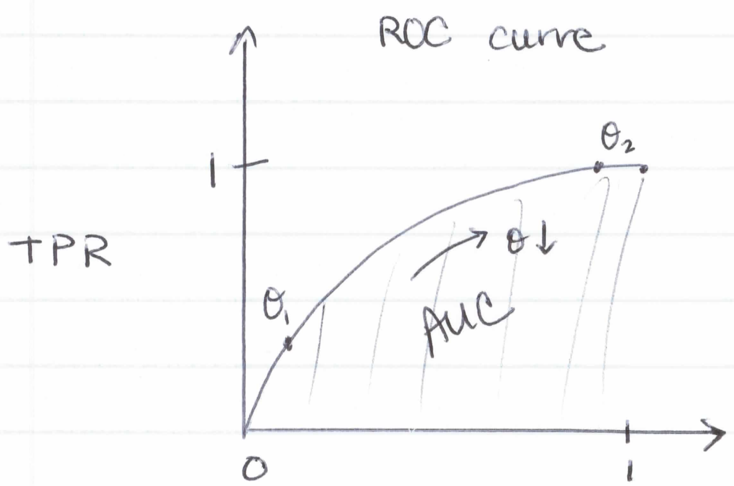
Q: How can we evaluate our classifier?

Actual \ Predicted	S	NS
S	TP	FN
NS	FP	TN

TP = true positive
TN = true negative
FP = false positive
FN = false negative

$$\text{Sensitivity} \triangleq TPR = \frac{TP}{TP + FN}$$

$$\text{Specificity} \triangleq TNR = \frac{TN}{FP + TN}$$



$$FPR = \frac{FP}{FP + TN}$$