TAM 598 Lecture 16:

Classification

Announcements:

- HW 4 covers lectures 13-16; due on Fri Apr 4

updated!

I. Classification / logistic regression

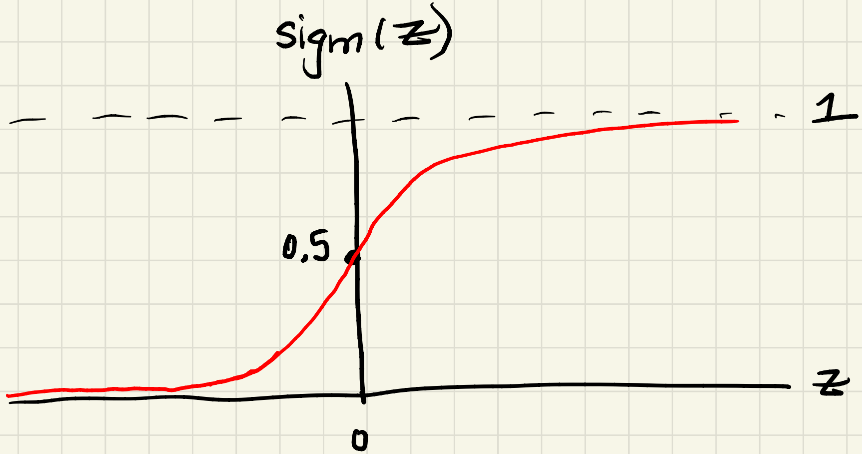
observations $\underset{\sim}{x}_{1:n} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$

targets $y_{1:n} = (y_1, \dots, y_n)$

consider binary classification, where $y = 0$ or $y = 1$.

probability that $y = 1$ conditioned on x:

where :



logistic regression $\rightarrow$ a generalized linear model pushed through a sigmoid function so it is mapped to [0, 1]

then

or

II. Likelihood of all observed data:

$$p(y_{1:n} \mid \underline{x}_{1:n}, \underline{w}) =$$

$$=$$

Then we obtain the best weight vector $\underline{w}$ using MLE:

$$\underline{w}^* =$$

$$=$$

Finding max $\underline{w}$ is equivalent to minimizing the loss

$$L(\underline{w}) = -\sum_{i=1}^{n} \left\{ y_i \,\mathrm{sigm}\left(\underline{w}^T \underline{\phi}(\underline{x}_i)\right) + (1-y_i)\left[1-\mathrm{sigm}\left(\underline{w}^T \underline{\phi}(\underline{x}_i)\right)\right] \right\}$$

this is called a cross-entropy loss function.
used also, eg, for DNNs that classify images.

III Making Decisions : say we have a point estimate $\underline{w}^*$

↳ $p(y \mid \underline{x}, \underline{w} = \underline{w}^*)$ is the prob. that $y=1$

↳ but you need to make a decision, $y=0$ or $y=1$

As before with decision making, need a loss function choice of cost function is subjective.

For binary classification, $\ell(\hat{y}, y)$ is a 2×2 matrix

Given $\ell(\hat{y}, y)$, your decision is the $\hat{y}$ that minimizes expected loss

IV. Diagnostics for Classification

- split dataset into training & validation
- two important metrics are

1) accuracy

2) confusion matrix

V. Multi-class classification

observations $\underset{\sim}{x}_{1:n} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$

targets $y_{1:n} = (y_1, \dots, y_n)$ ← discrete labels

Now we have K possible values for labels so