(d) methods:

For NB: We can use either MLE or MAP.

For LR: Choose parameters $W = \langle W_0, \dots, W_n \rangle$ to maximize conditional likelihood of training data. called MCLE

Since there is no closed form solution to maximizing $l(W)$

Details are in the following: we use ~~the~~ gradient ascent.

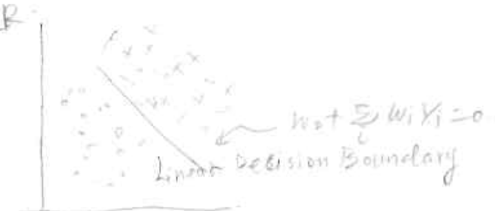
2. a) NB is classed a generative classifier, because we can view the distribution $P(X|Y)$ as describing how to generate random instances X conditioned on the target attribute Y.

LR is referred to as a discriminative classifier because we can view the distribution $P(Y|X)$ as directly discriminating the value of the target value Y for any given instance X.

b)

μ: decide the position of the contour

Σ: decide the shape of the contour.



$W = \langle W_0, \dots, W_n \rangle$ is the parameters learned.

(d) of 1.

d. methods to do MLE.

For NB:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} P(D|\theta)$$
$$= \arg\max_{\theta} \prod_{i=1}^{n} P(Y_i|\theta)$$

Then take derivative and set it to 0, then we get $\hat{\theta}_{MLE}$.

$$\hat{\theta}_{MAP} = \arg\max_{\theta} P(\theta|D)$$
$$= \arg\max_{\theta} P(D|\theta) \cdot P(\theta)$$

Then take derivative and set it to 0, then we get $\hat{\theta}_{MAP}$

For LR: we solved for LR parameters with MCLE

$$l(w) = \log \prod_{j=1}^{n} P(y^{(j)} = y \mid X^{(j)} = x; w)$$

~~Since~~ Since there is no closed-form solution to maximizing $l(w)$.

We use gradient ascent.