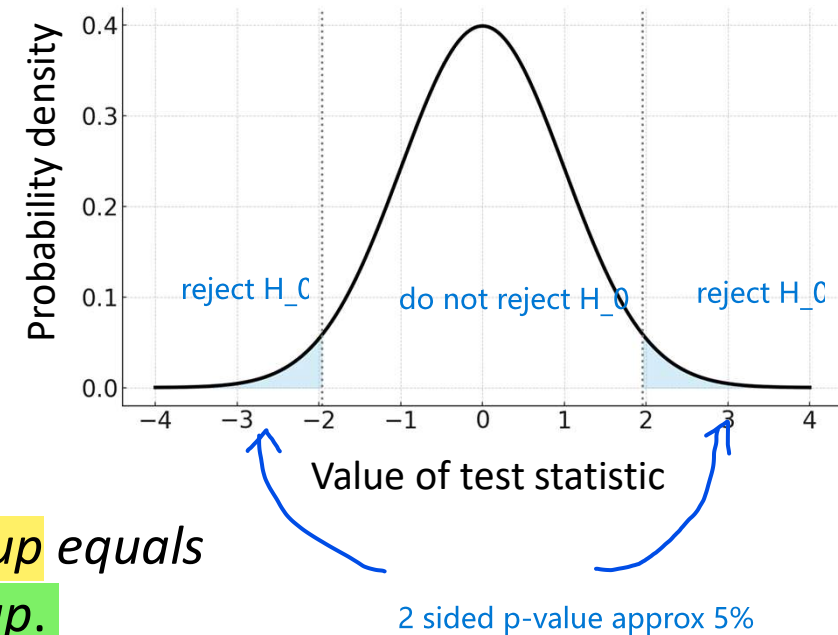


Hypothesis testing and the p-value

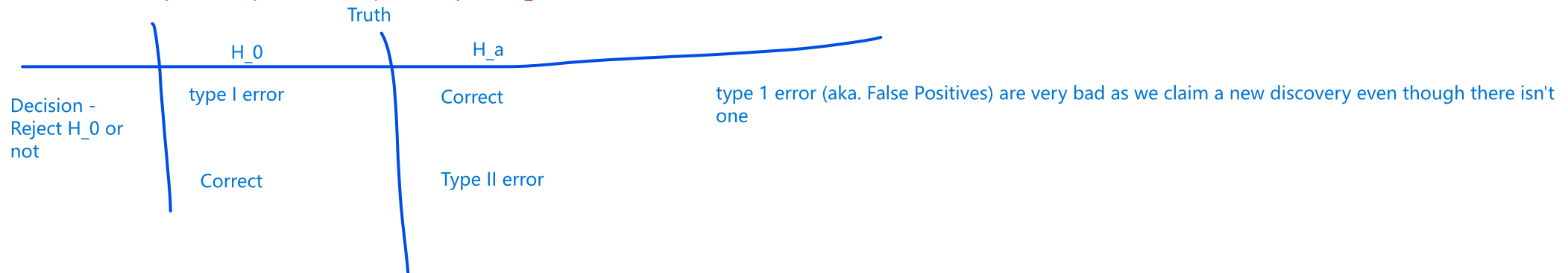
Example: Imagine we have developed a drug to treat blood pressure. Of course, we would like to discover that there is a difference between the mean blood pressure in the two groups. We construct a null hypothesis corresponding to no difference.

H_0 : the **expected blood pressure of mice in the control group** equals the **expected blood pressure of mice in the treatment group**.



p-value is defined as probability of observing a test statistic equal to or more extreme than the observed statistic, under the assumption that H_0 is in fact true. T small p-value provides evidence against H_0 .

It is NOT correct to say that the p-value is the probability that H_0 is



Classification tasks

<https://teachablemachine.withgoogle.com/>

Teachable Machine

Train a computer to recognize your own images, sounds, & poses.

A fast, easy way to create machine learning models for your sites, apps, and more – no expertise or coding required.

Get Started



Teachable Machine is flexible – use files or capture examples live. It's respectful of the way you work. You can even choose to use it entirely on-device, without any webcam or microphone data leaving your computer.



Images

Teach a model to classify images using files or your webcam.



Sounds

Teach a model to classify audio by recording short sound samples.

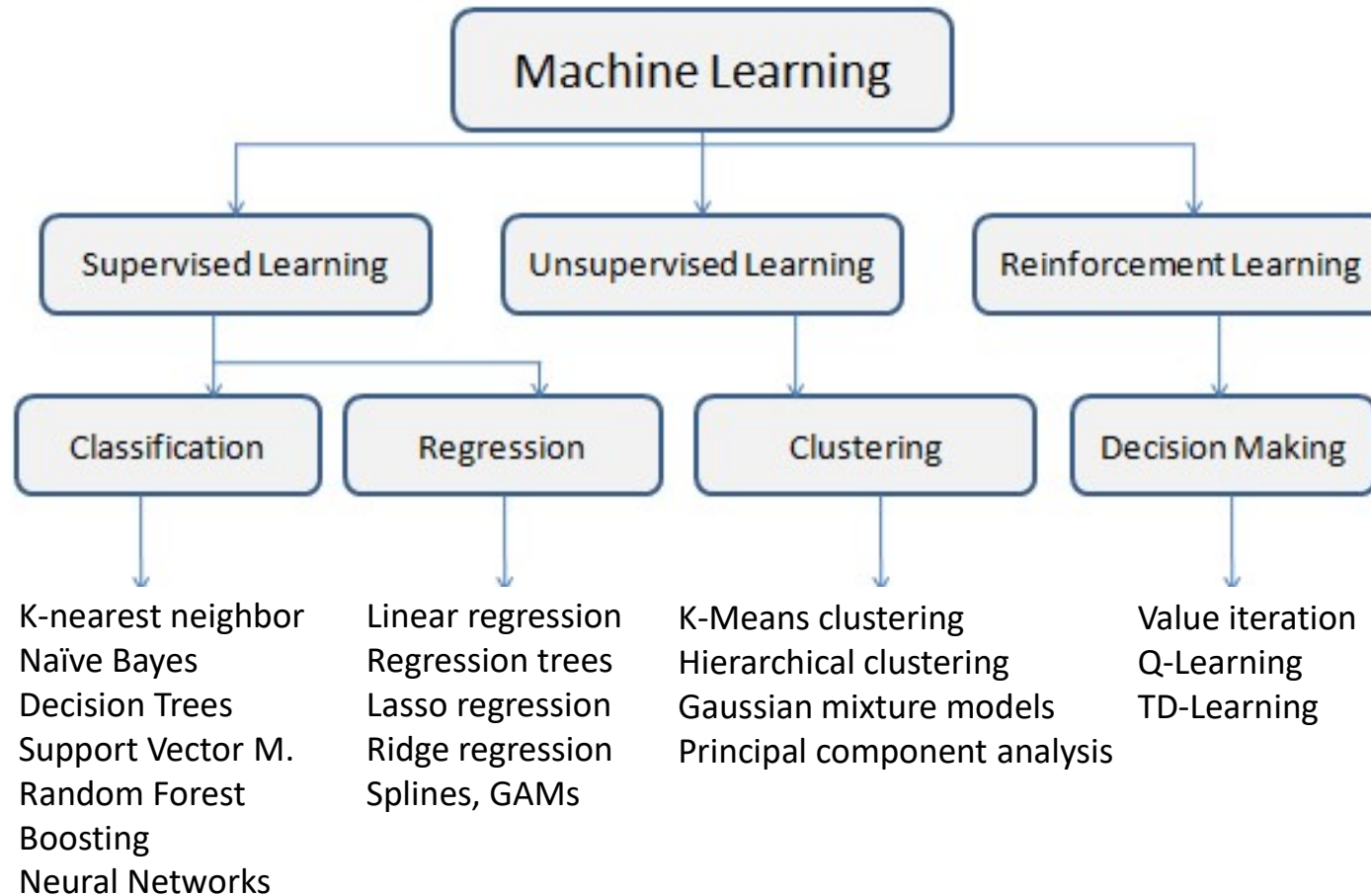


Poses

Teach a model to classify body positions using files or striking poses in your webcam.

Classification Algorithms: Naïve Bayes and Logistic Regression

Possible categorization of Machine Learning



There is no free lunch theorem: Just because there is a certain tool to solve one learning problem, doesn't mean that it can solve a different problem as well.

Can we use Linear Regression for Classification tasks?

Example: Want to investigate whether students can pay their bills based on their bank account balance.

Can we use Linear regression (aka least square) for classification? - Yes
for example: $y = 0$ if no or 1 if yes

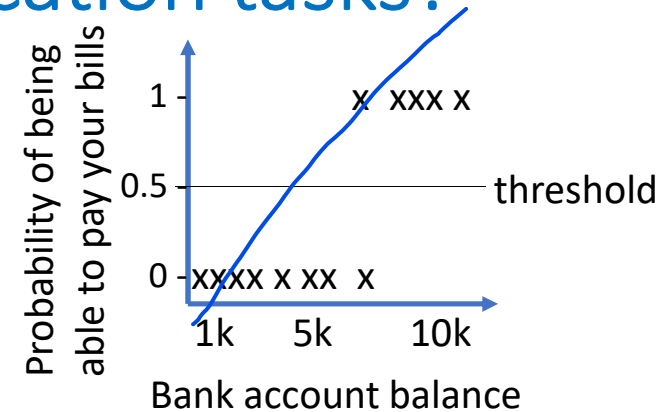
can use linear regression of y on x and classify as Yes, if \hat{y} greater than 0.5

Disadvantage: if we take y as probability, then it could happen that linear regression will predict probability outside of $[0:1]$, which makes them hard to interpret as probabilities

Nevertheless, the predictions provide an ordering and can be interpreted as a crude probability estimates.

Classification of linear regression to predict a binary response will be the same as for the linear discriminant analysis (LI

Linear regression shouldn't be used if the response variable has 3 or more variables



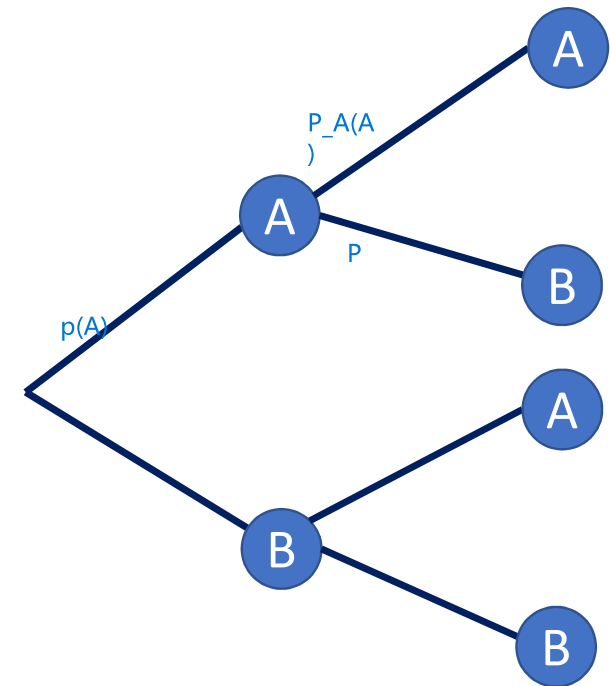
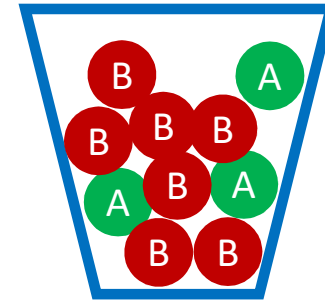
Quiz question - Bayes' theorem

Imagine we have a bucket with 3 balls with letter A and 7 balls with letter B. We are now withdrawing one ball and then a second ball without putting the first one back.

Which statements for the probability, P, are correct?

- a) $P(A) = 3/10$
- b) $P(B|A) = 7/9$
- c) $P(A) P(B|A) = P(B) P(A|B)$
- d) Bayes' theorem states:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$



The Naïve Bayes Classifier

use Bayes theorem when we want to predict class Y given the feature set X, yields: $\underbrace{\Pr(Y = k | X = X_0)}_{\text{posterior Pr}} = \frac{\overbrace{\Pr(Y=k)}^{\text{Prior Pr}} \cdot \overbrace{\Pr(X = x_0 | y=k)}^{\text{likelihood}}}{\underbrace{\Pr(X = X_0)}_{\text{evidence}}}$

assumption: complete independence between the features (aka. the covariance matrix is diagonal)

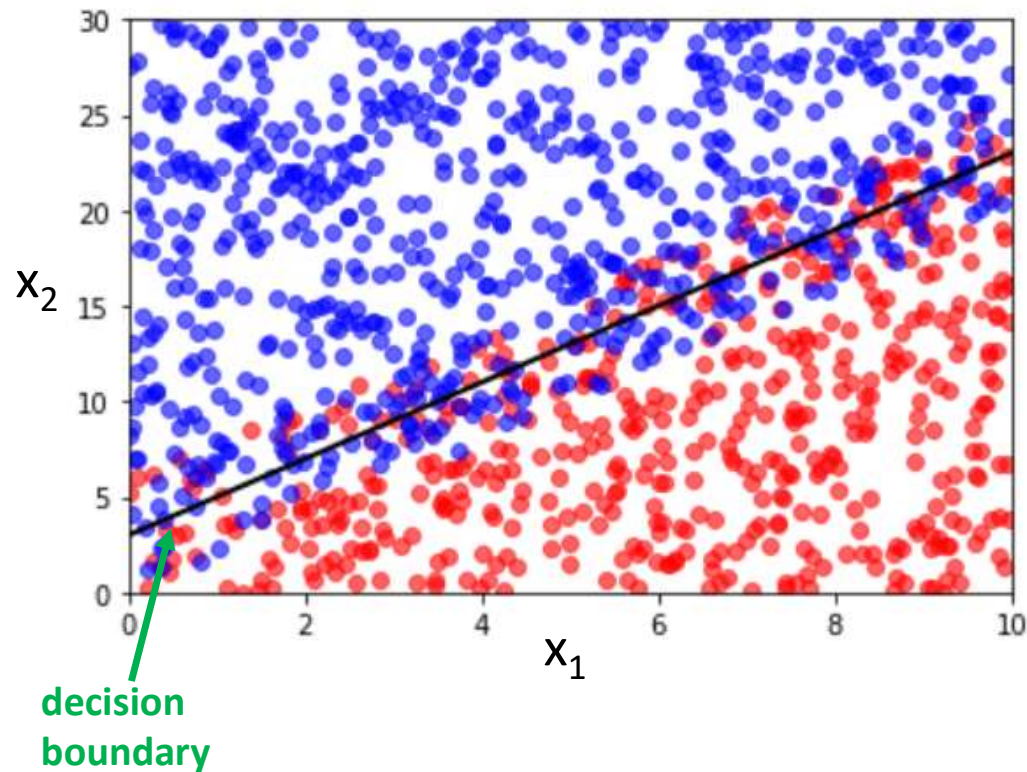
The Bayes classifier assumes that the test error rate is minimized by a simple classifier that assigns each observation to the most likely class given its predictor values

simply assign test observation with predictor vector X_0 to class k for which the conditional probability $\Pr(Y = k | X = X_0)$ is largest.

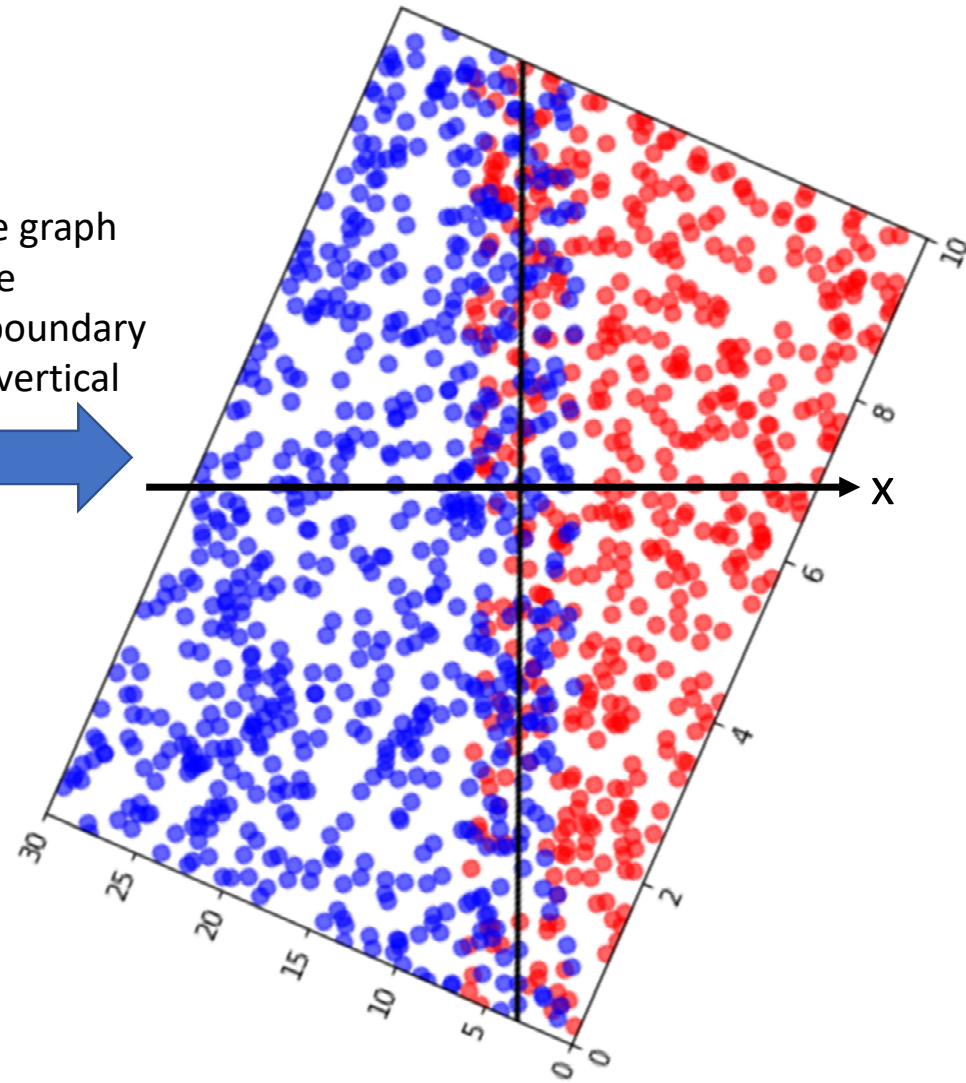
Bayes classifier produces the lowest possible test error rate, called the Bayes error rate analogous to irreducible error
Unattainable gold standard because we don't know the conditional distribution of Y given X
Therefore, computing bayes classifier is impossible.

Towards logistic regression

probability to find a red data point

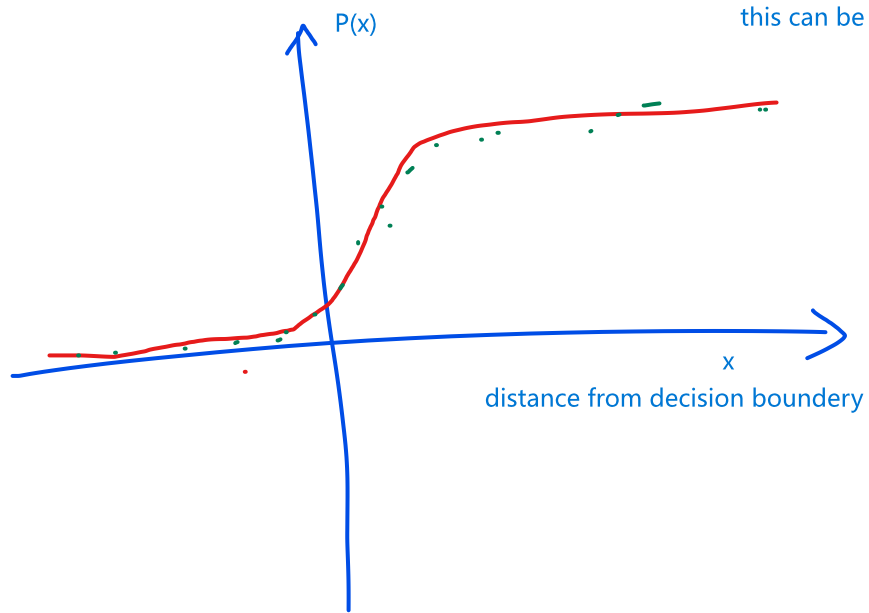


Rotate the graph so that the decision boundary becomes vertical



Jupyter Notebook: LinReg_LogReg_Python.ipynb

Logistic Regression



this can be approximated by sigmoid function: $\sigma(x) = 1 / (1 + e^{-x})$

Linear regression: $Y = \text{Beta}_0 + \text{Beta}_1 \cdot X$

Logistics regression: $P(x) = \sigma(\text{Beta}_0 + \text{Beta}_1 \cdot x)$

Quiz question

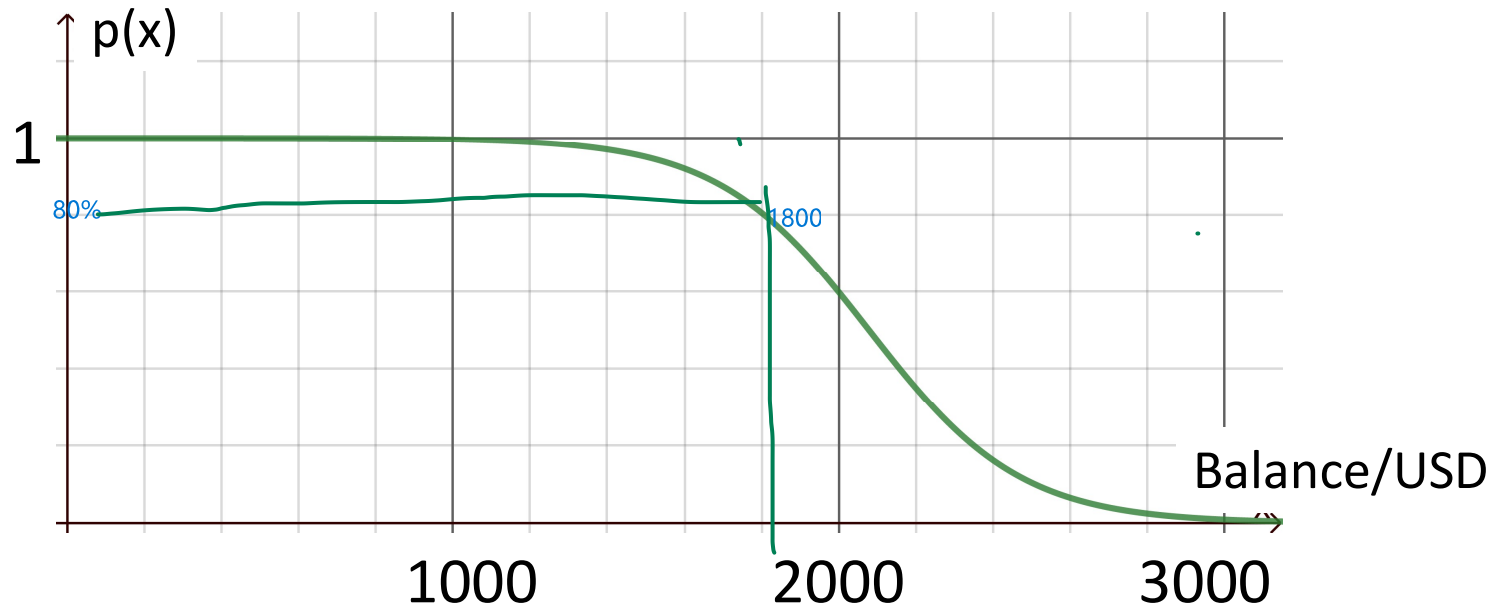
The probability of not being able to pay the bills $p(x)$ was estimated through the account balance and yielded $\hat{\beta}_0 = 10.6$ and $\hat{\beta}_1 = -5.1$.

The probability

$$p(X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}}$$

$$= \frac{1}{1+e^{-(10.6-5.1X)}}$$

was then plotted.



True or false: The probability of not being able to pay the bills is approx. 80% for a balance of 1800 USD.

Logistic regression and maximum likelihood

what are the odds that the event occurs? - it is the (probability/ counter probability) = $p(x)/(1-p(x))$

ex: throw a 6 with a regular dice, so we have:

odd = $(1/6):(5/6) = 1/5$

Let's calculate the natural logarithm of the odds, $\log(\text{odds})$ also known as Logits (thus the name logistic regression):

$$\ln(p(x)/(1-p(x))) = \ln (1/ (p(x) ^{-1} -1)) =$$

Using now the method of max likelihood to estimate the parameters. We choose Beta1 & Beta0 in a way that maximize the likelihood of the observed data

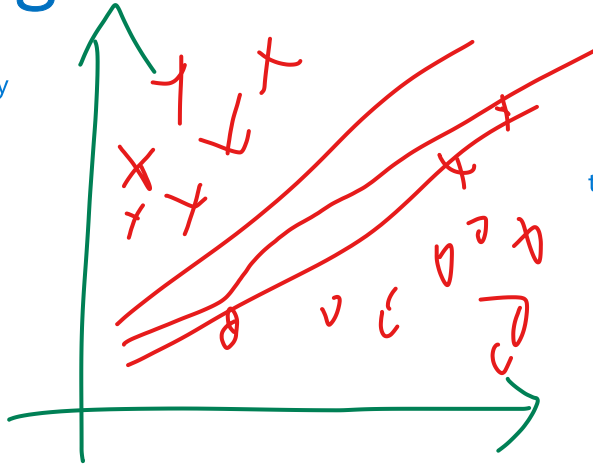
Likelihood function (Beta0, Beta1)= multiplication of $p(x)$ * multiplication $(1-p(x))$

The likelihood gives the probability of the observed zeros and ones

Weakness of logistic regression

well separated data makes trouble for logistics regression to classify
no guarantee that this is pitched

Logistics regression can give poor result
when two classes are perfectly separated by linear decision
bounder
Because $L(\beta_0, \beta_1)$ cannot be maximized as there is always
something bigger
thus $\hat{\beta}_0$ and $\hat{\beta}_1$ are undefined



test sample would get classified as X for this decision boundary

Multiclass logistic regression

with p variable we get $\ln(p(x)/(1-p(x))) = \text{Beta}_0 + \text{Beta}_1X_1 + \dots + \text{Beta}_pX_p$

multi class logistics regression aka multinomial regression uses a linear function for each class K:

$$\Pr(y=k|X) = \frac{e^{(\text{Beta}_0 + \text{Beta}_1X_1 + \dots + \text{Beta}_pX_p)}}{\sum (e^{\text{Beta}_0 + \text{Beta}_1X_1 + \dots + \text{Beta}_pX_p})}$$

data augmentation ?

Case-control sampling

especislly in medical studies we have extremely imbalanced ckass sizes, for eg the prior probabilities of classes are very unequal

ex: assume the overall probability of a heartattach is $\pi = 0.05$ in a certain country

in our medical study, however we have 160 cases of heart attach and 302 controls.

thus the probability in this sample is $\pi = 0.35 = \text{cases} / (\text{controls} + \text{cases})$

There is an imbalance as we didn't do proportional sampling of people with and without heartattack. So we can simple shrift the intercept of this imbalance using the following transformation:

$$\text{trasnformed_Beta0}^{\wedge} = \text{Beta0}^{\wedge} + \ln(\pi / 1-\pi) - \ln(\pi / 1-\pi) = \text{Beta0}^{\wedge} - 2.94 + 0.62 = \text{Beta0}^{\wedge} - 2.32$$