3D geoinformation
Department of Urbanism
Faculty of Architecture and the Built Environment
Delft University of Technology

GEO5017
Machine Learning for the Built Environment

Lecture
Classification

Shenglan Du

Today's Agenda

- Previous Lecture: Supervised Learning

- Bayes Classification
  - Probability Basics
  - Bayes Classifier

- Linear Classification
  - Fisher Classifier
  - Logistic Classifier

Today’s Agenda

- Previous Lecture: Supervised Learning 👈

- Bayes Classification
  - Probability Basics
  - Bayes Classifier

- Linear Classification
  - Fisher Classifier
  - Logistic Classifier

Supervised Learning

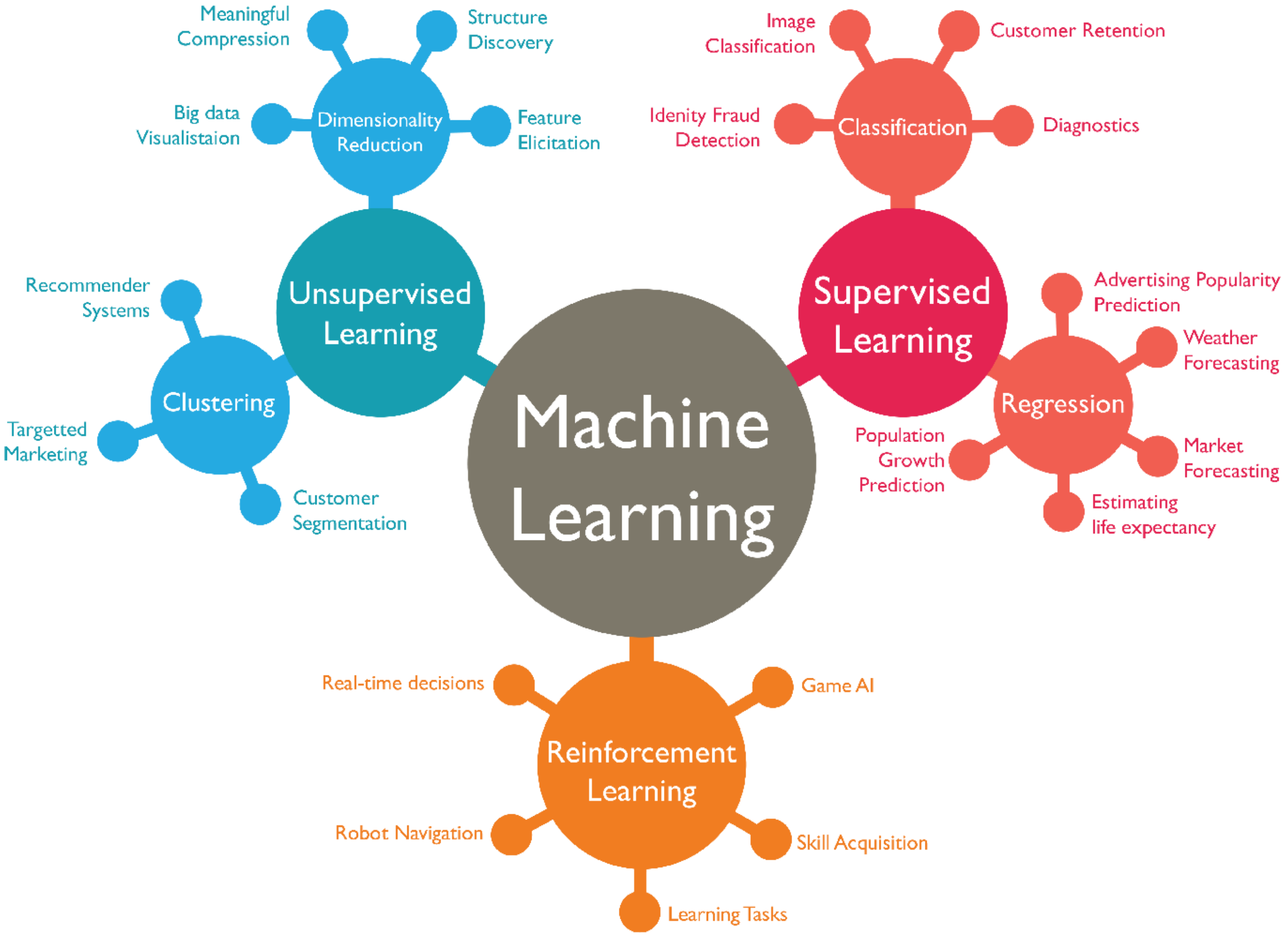
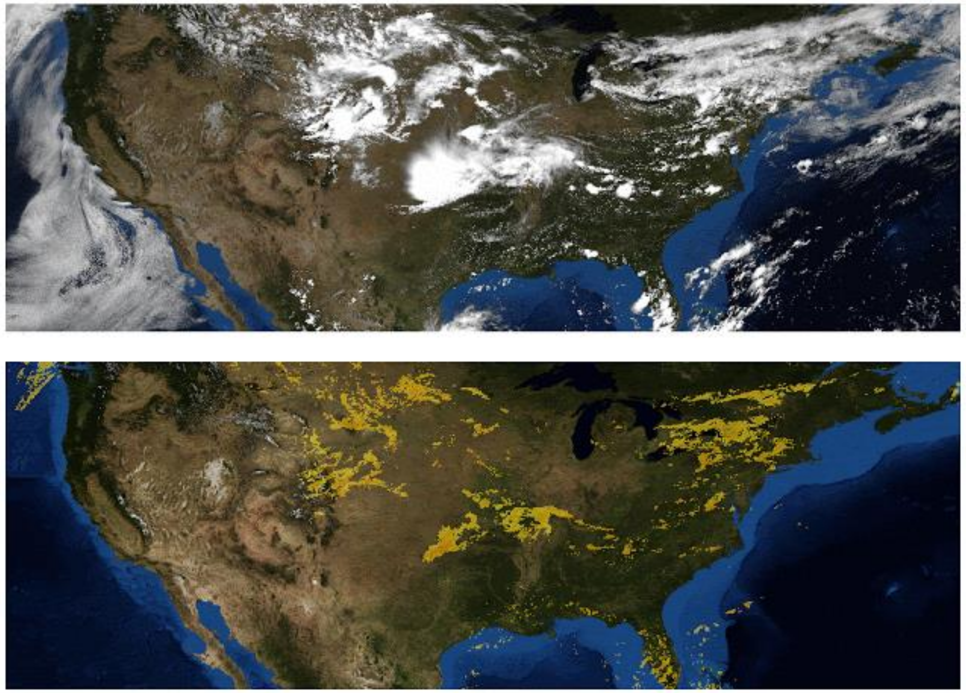


Image source: https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications

Supervised Learning

- An example: weather forecasting



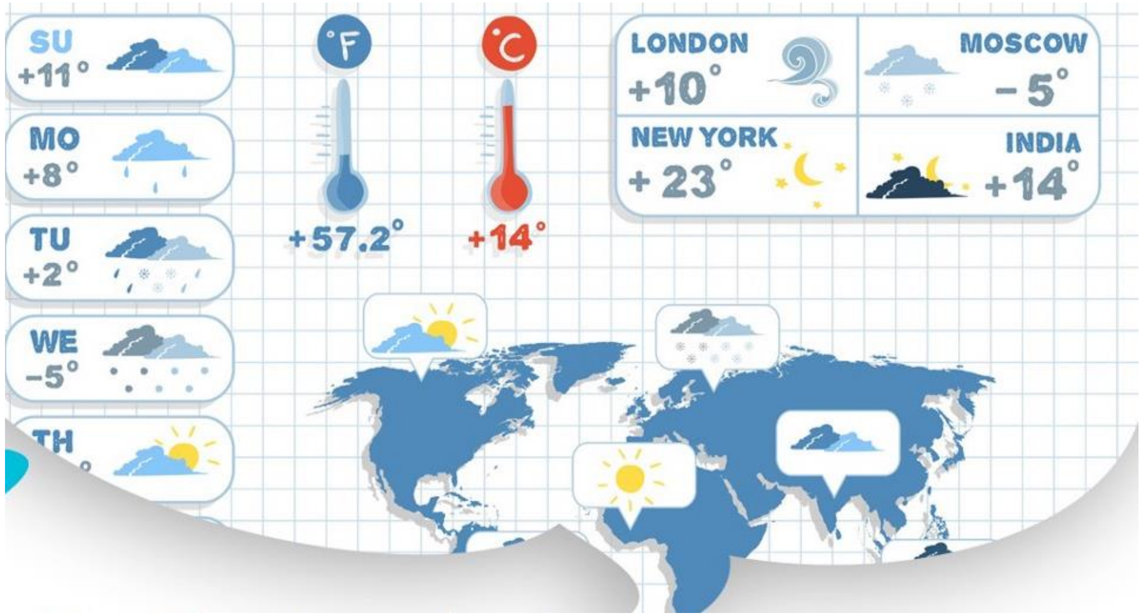


Image source: https://medium.com/@chirag071857/machine-learning-in-weather-forecasting-b47ba8edd694

Supervised Learning

- An example: image analysis

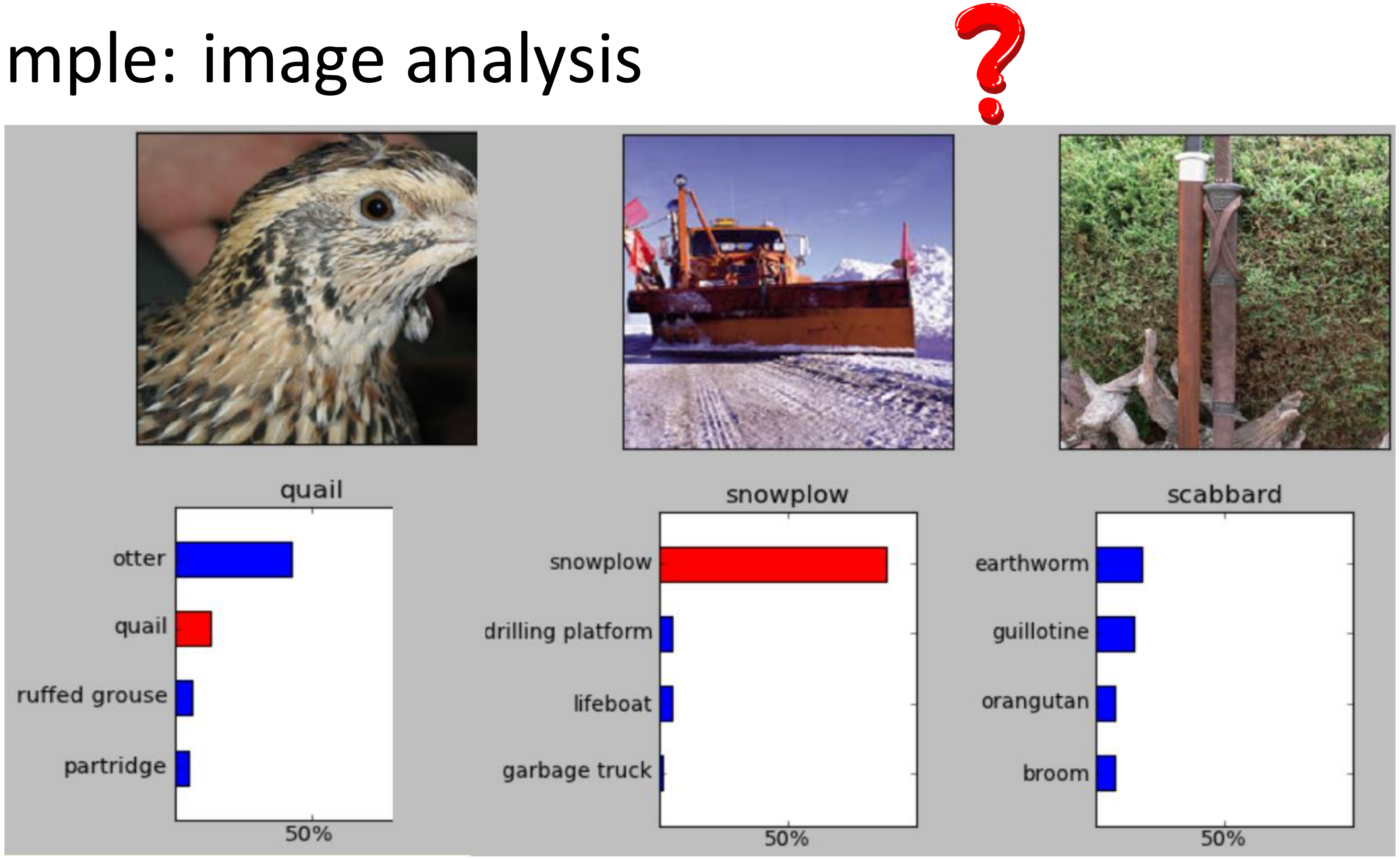


Image source: https://www.cs.toronto.edu/~hinton/coursera_slides.html

Supervised Learning: Classification

- We usually have a set of input data represented as feature vectors:

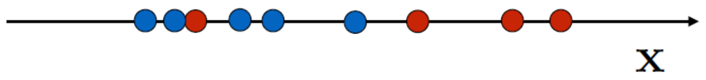
$$\boldsymbol{x} = (x_1, x_2, x_3 \dots x_p)^T$$

- Classification aims to specify which category/class $\boldsymbol{y}$ some input data $\boldsymbol{x}$ belong to

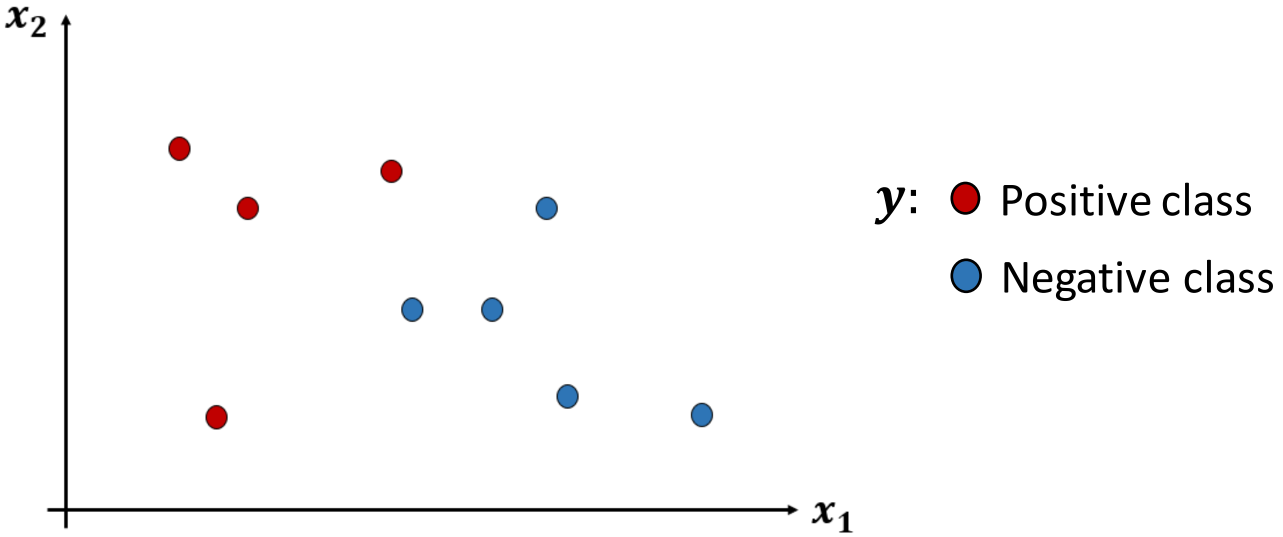
Supervised Learning: Classification

$$\boldsymbol{x} = (x_1, x_2, x_3 \ldots x_p)^T$$

- P indicates the feature space dimension:
  - 1D feature space:

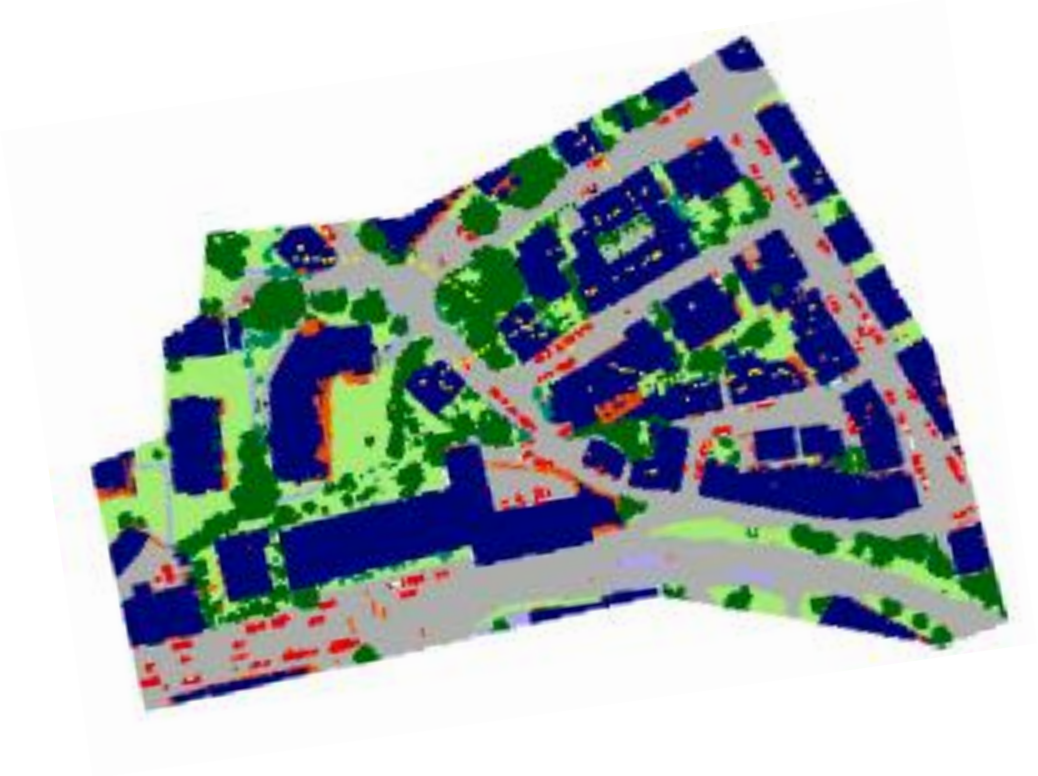


  - 2D feature space:



Supervised Learning: Classification

- An example of point cloud semantic classification



$\boldsymbol{x} = (x, y, z, r, g, b, intensity \ldots)^T$

$\boldsymbol{y}$:
- High vegetation
- Low vegetation
- Building
- Road
- Grass land

Image source: https://www.sciencedirect.com/science/article/pii/S0924271620300605

Supervised Learning: Classification

- Two classification approaches:
  - **Generative approach**: model the probability distribution of feature $\boldsymbol{x}$ and label $\boldsymbol{y}$
    - Bayes classifier
    - Gaussian mixture model
  - **Discriminant functions**: model a function that directly map from feature $\boldsymbol{x}$ to label $\boldsymbol{y}$
    - Linear classifier (Fisher, Logistic regression, SVM)
    - Non-linear classifier (Decision tree, Neural networks)

Today's Agenda

- Previous Lecture: Supervised Learning
- Bayes Classification 👈
  - Probability Basics
  - Bayes Classifier
- Linear Classification
  - Fisher Classifier
  - Logistic Classifier

Bayes Classification

- A simple scenario: A tree or a building?

Image source 1: https://en.wikipedia.org/wiki/Tree#/media/File:Ash_Tree_-_geograph.org.uk_-_590710.jpg

Image source 2: https://en.wikipedia.org/wiki/Wilder_Building#/media/File:WilderBuildingSummerSolstice.jpg

Bayes Classification

- A simple scenario:
  - Buildings have planar surfaces
  - Trees have noisy, near round surfaces

- The machine detected that the input object has planar surfaces. What the object do you guess to be?

Bayes Classification

- It’s very likely to be a building!

- But how do machines interpretate the word “likely”?

Today's Agenda

- Previous Lecture: Supervised Learning
- Bayes Classification
  - Probability Basics
  - Bayes Classifier
- Linear Classification
  - Fisher Classifier
  - Logistic Classifier

Probability Basics

- Product rule:

$$P(X,Y) = P(X)\,P(Y|X)$$

- Bayes rule:

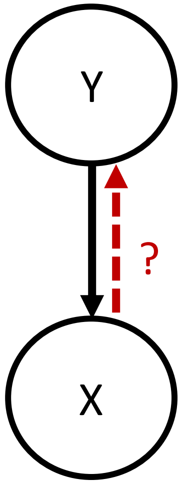
$$P(Y)\,P(X|Y) = P(X)\,P(Y|X)$$

$$P(Y|X) = \frac{P(Y)\,P(X|Y)}{P(X)}$$

Probability Basics

- Given feature $\boldsymbol{x}$ and label y

$$P(y|\boldsymbol{x}) = \frac{P(y)\,P(\boldsymbol{x}|y)}{P(\boldsymbol{x})}$$



  - $P(\boldsymbol{x}|y)$: class conditional probability
  - $P(y)$: class prior probability
  - $P(y|\boldsymbol{x})$: class posterior probability

Probability Basics

- Assume equal priors for both buildings and trees

$$P(y = b) = P(y = t) = 0.5$$

Probability Basics

- Assume we have the class conditional probabilities as follows

$$P(x = planar|y = b) = 0.8$$
$$P(x = round|y = b) = 0.2$$
$$P(x = planar|y = t) = 0.25$$
$$P(x = round|y = t) = 0.75$$

Probability Basics

- building:

$$P(y = b|x = planar) = \frac{P(y = b)\, P(x = planar|y = b)}{P(x = planar)}$$

$$= \frac{0.5 * 0.8}{P(x = planar)}$$

- tree:

$$P(y = t|x = planar) = \frac{P(y = t)\, P(x = planar|y = t)}{P(x = planar)}$$

$$= \frac{0.5 * 0.25}{P(x = planar)}$$

Probability Basics

- Prior:

$$P(y = b) = P(y = t)$$

- Posterior:

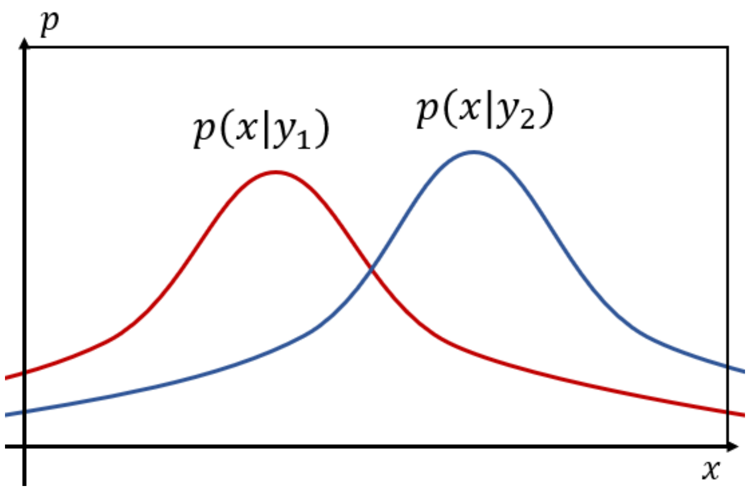
$$P(y = t|x = planar) \ll P(y = b|x = planar)$$

Today's Agenda

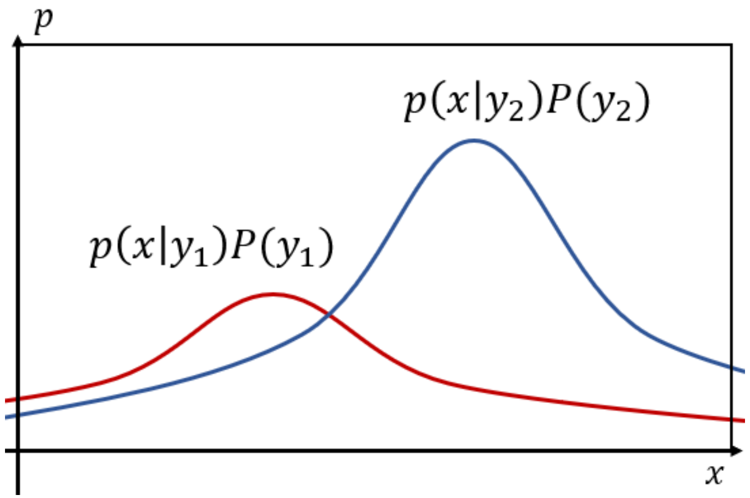
- Previous Lecture: Supervised Learning
- Bayes Classification
  - Probability Basics
  - Bayes Classifier
- Linear Classification
  - Fisher Classifier
  - Logistic Classifier

Bayes Classifier

- Step 1: estimate the class conditional probabilities



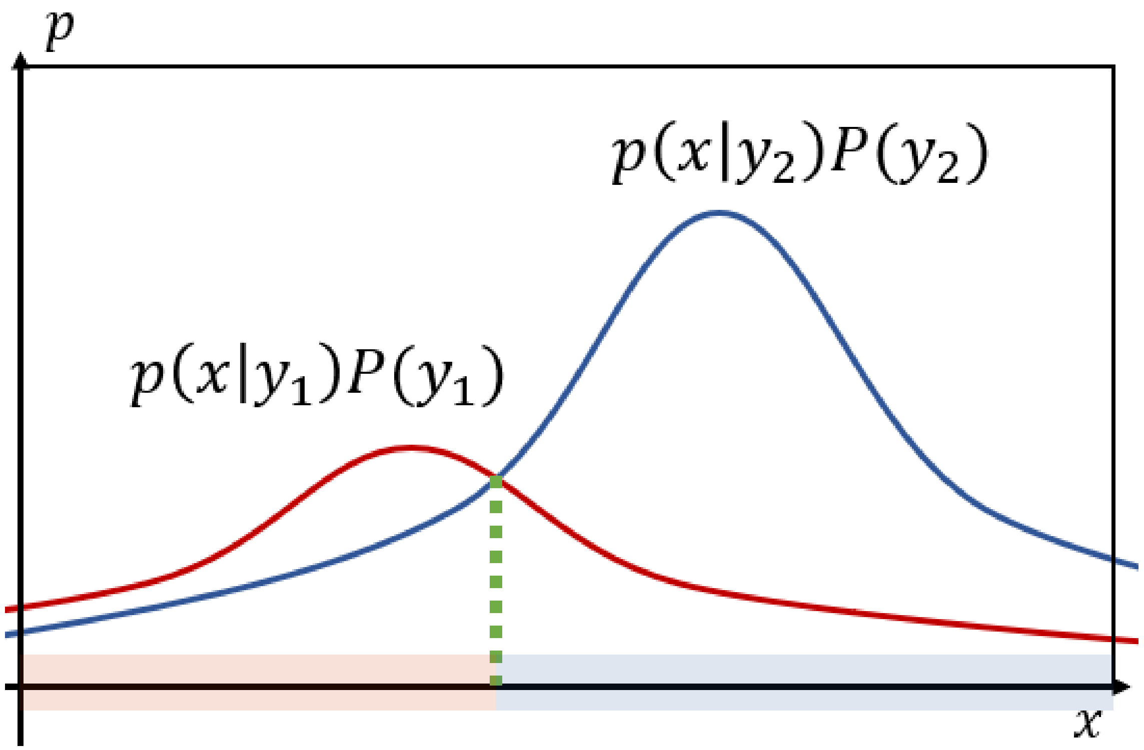
- Step 2: multiply with class priors



- Step 3: compute the class posterior probabilities

Bayes Classifier

- Step 4: find the classification boundary



Bayes Classifier

- The Bayes rule provides an approach of describing the uncertainty quantitatively, allowing for the optimal prediction given the observations present

- Bayes serves as the foundation for the modern machine learning

Bayes Error

- All models are wrong but some are useful… --LeCun

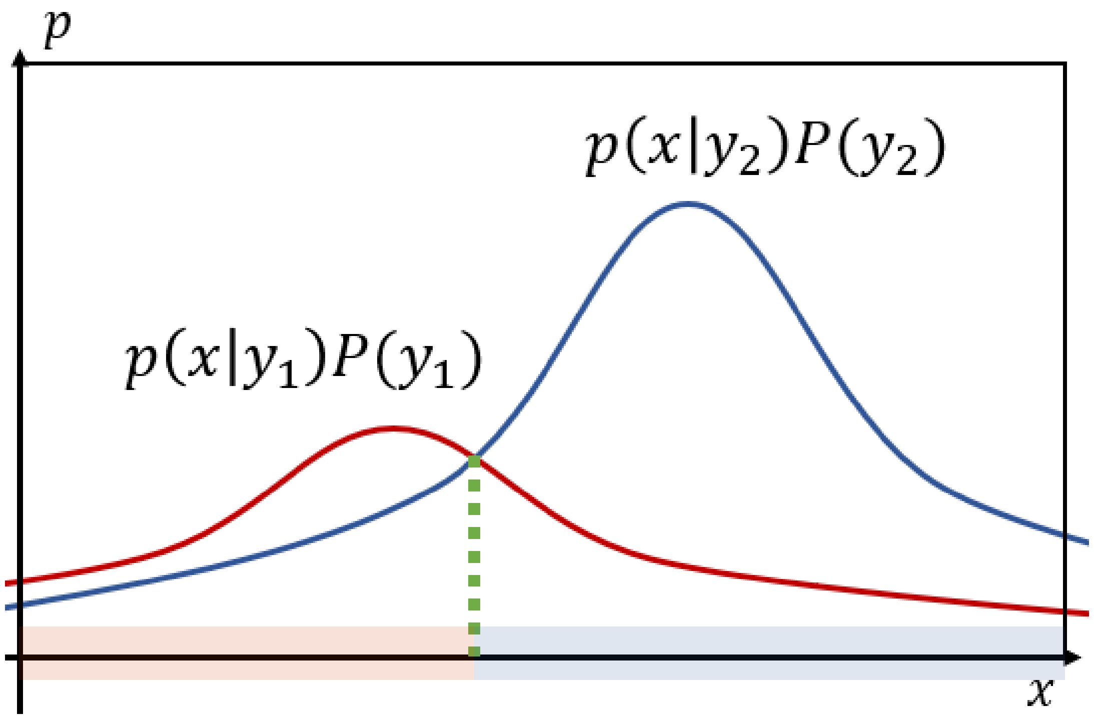


Image source: Wikipedia

Bayes Error

- Where is the error?
  - All trees have spherical surfaces
  - All buildings have cube-shapes
  - All rabbits have long ears
  - All Scotland sheeps are black
  - ……

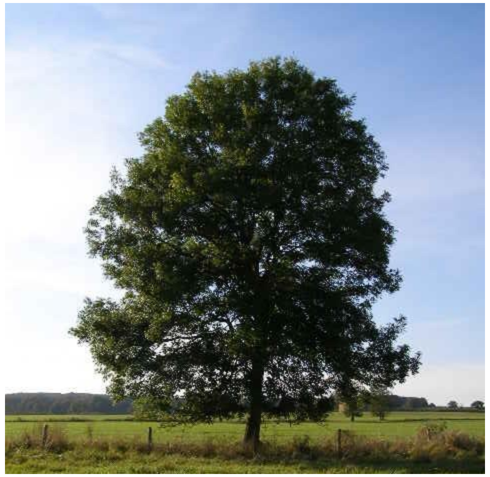



Image source: Google

Bayes Error

- Where is the error?

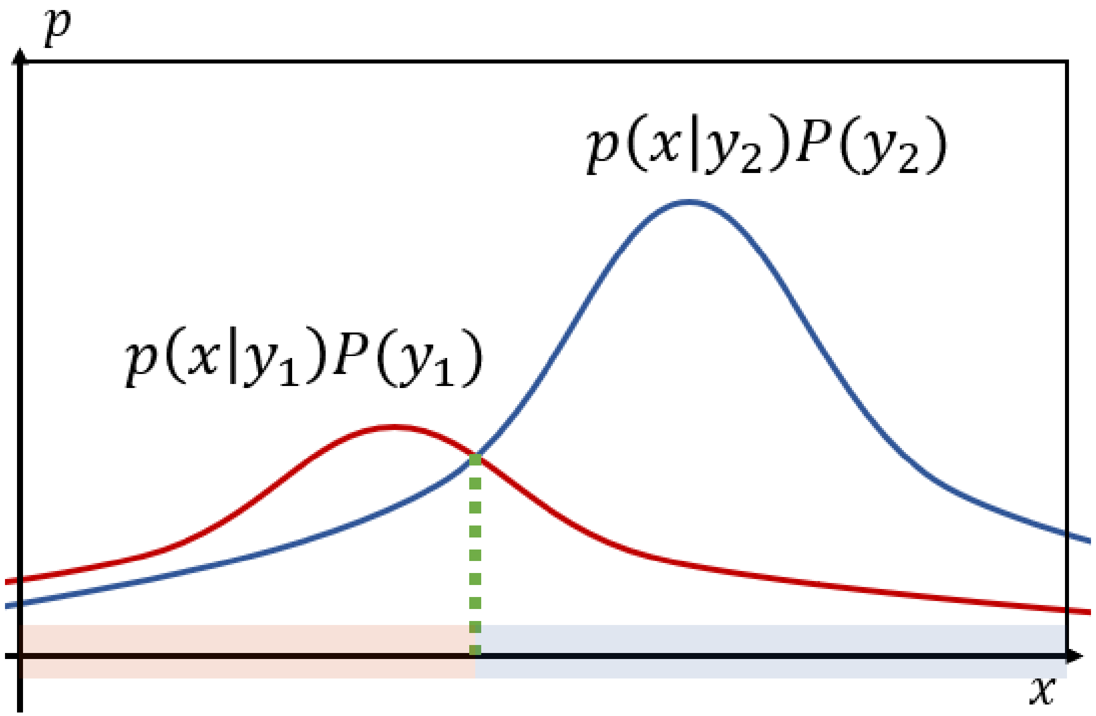
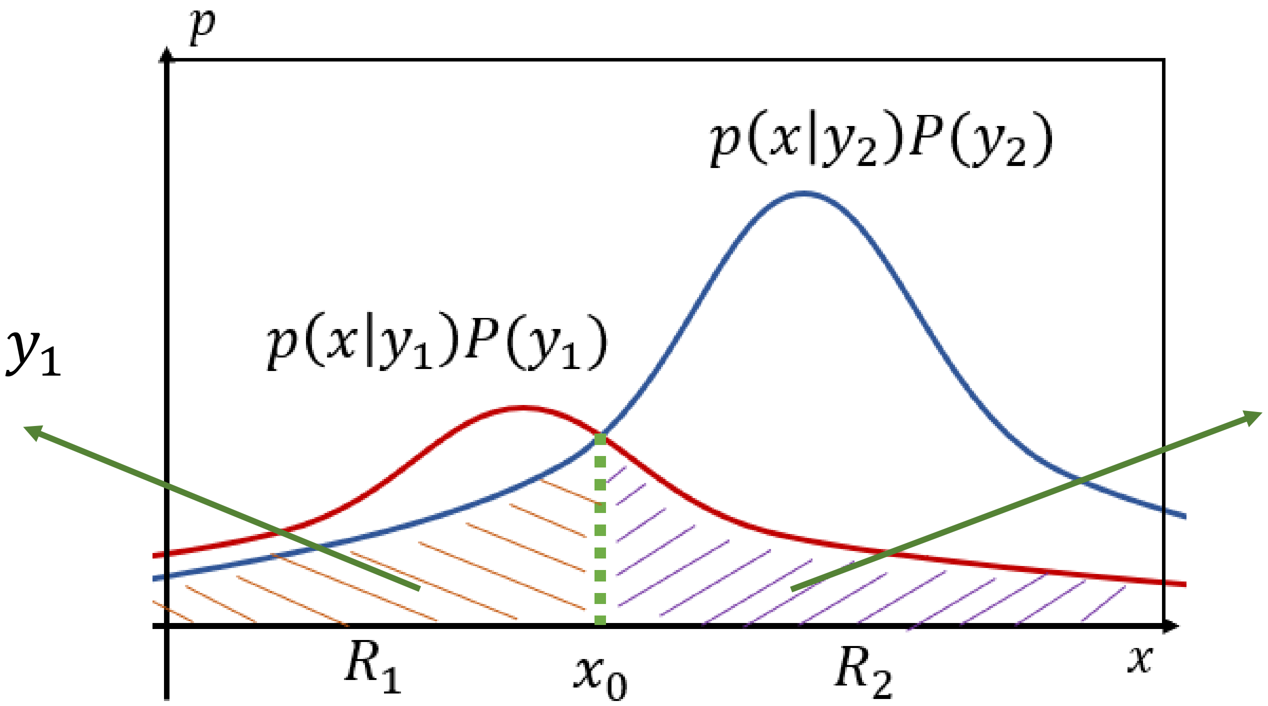


Image source: Wikipedia

Bayes Error

Misclassifying y_2 to y_1



Misclassifying y_1 to y_2

$$P(e) = \int_{-\infty}^{x_0} p(x|y_2)P(y_2) + \int_{x_0}^{\infty} p(x|y_1)P(y_1)$$

Bayes Error

- It’s the minimum attainable error using any kinds of existing models (SVM, RF, Neural networks)

- It doesn’t depend on the ML model that you apply, but only on the data distribution

- We cannot obtain it as we don’t have true distributions of real world

Minimizing the Risk

- Healthy or ill?
  - Assigning healthy to ill will cause panic to the patient
  - Assigning ill to healthy has more severe outcome

Image source: Charles Choi. Medical Imaging AI Software Is Vulnerableto Covert Attacks. IEEE Spectrum. 04 Jun 2018

Minimizing the Risk

- Assume: $y_1 = healthy,\ y_2 = ill,\ \ \lambda_{ij}$ is the cost of assign class i to j

- Classifying with risk we have:
  - Assign **x** to y_1 if

$$\lambda_{21} p(\boldsymbol{x}|y_2) P(y_2) < \lambda_{12} p(\boldsymbol{x}|y_1) P(y_1)$$

  - Assign **x** to y_2 otherwise

Image source: Charles Choi. Medical Imaging AI Software Is Vulnerableto Covert Attacks. IEEE Spectrum. 04 Jun 2018

Today's Agenda

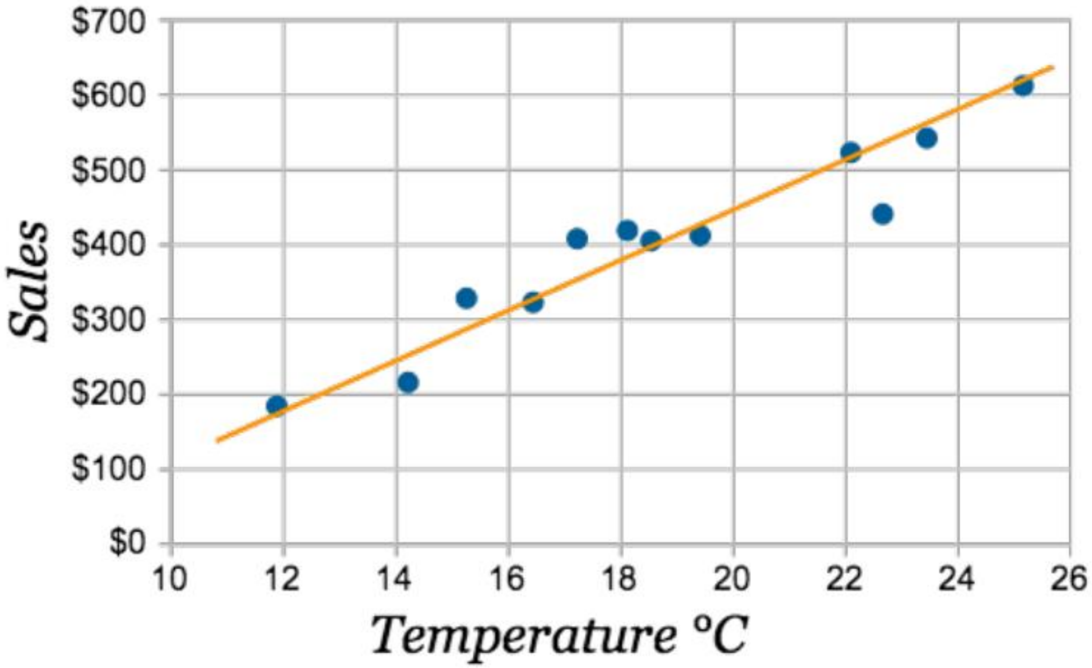
- Previous Lecture: Supervised Learning

- Bayes Classification
  - Probability Basics
  - Bayes Classifier

- Linear Classification 👈
  - Fisher Classifier
  - Logistic Classifier

Linear Classification

- Review Linear Regression:



Linear Classification

- Review Linear Regression:

$$y_i = \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{i}} + b$$

  - $\boldsymbol{x}_{\boldsymbol{i}} = (x_1, x_2, x_3 \dots x_p)^T$
  - $\boldsymbol{w} = (w_1, w_2, w_3 \dots w_p)$
  - b is the bias scalar value
  - y_i is the output scalar which should be continuous in its domain

Linear Classification

- Review Linear Regression:

$$y_i = \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{i}} + b$$

  - Parameters are determined by minimizing the square of errors
  - Optimization is achieved by gradient descent
  - A close form solution can be found:

$$(X^T X)^{-1} X^T Y$$

Linear Classification

- We consider constructing linear functions of input $\boldsymbol{x}$ to describe decision boundaries

$$y = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + b$$

  - $y = const$ determines a decision boundary
  - A decision boundary is a (D-1) dimension hyperplane of D dimension input feature space

Today’s Agenda

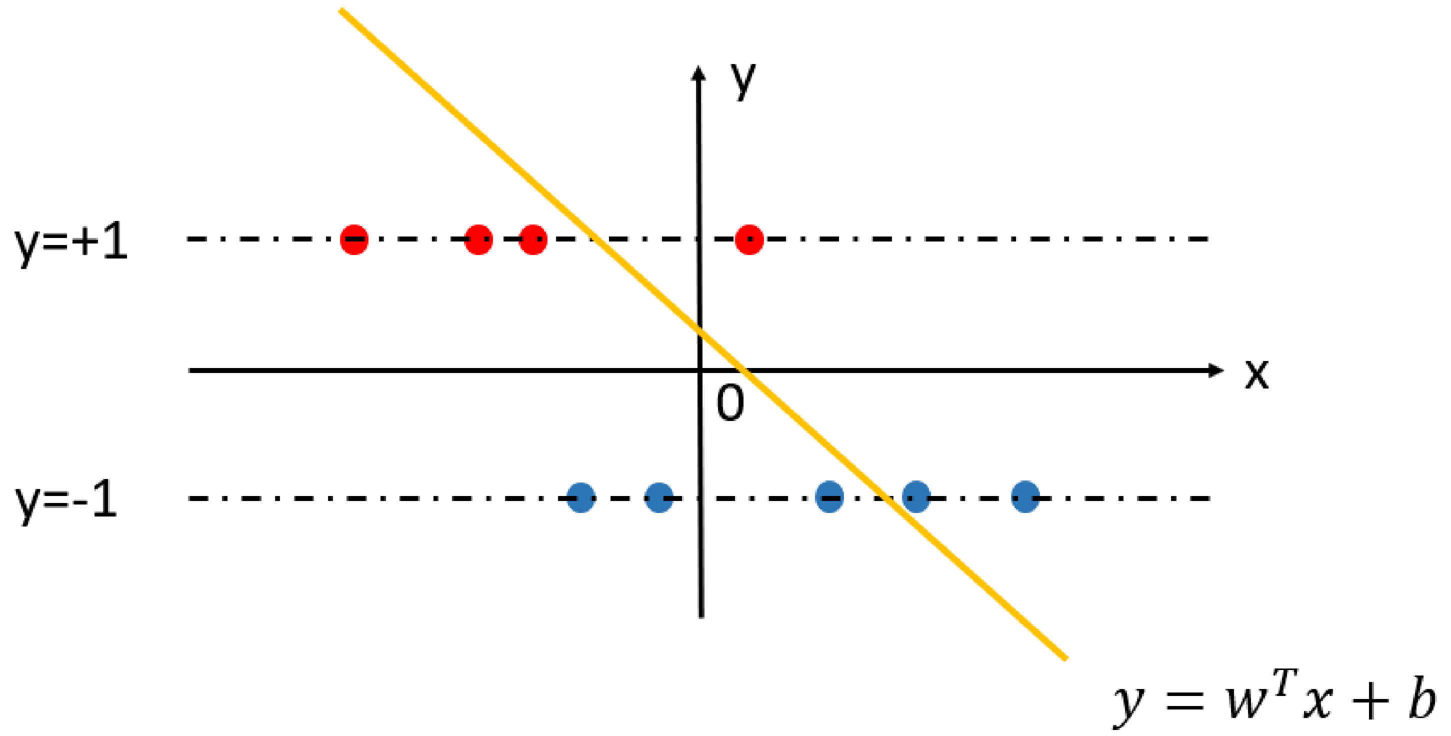
- Previous Lecture: Supervised Learning

- Bayes Classification
  - Probability Basics
  - Bayes Classifier

- Linear Classification
  - Fisher Classifier 👉
  - Logistic Classifier

Fisher Classifier

- Also refers to the standard linear classifier in the context of this course



Fisher Classifier

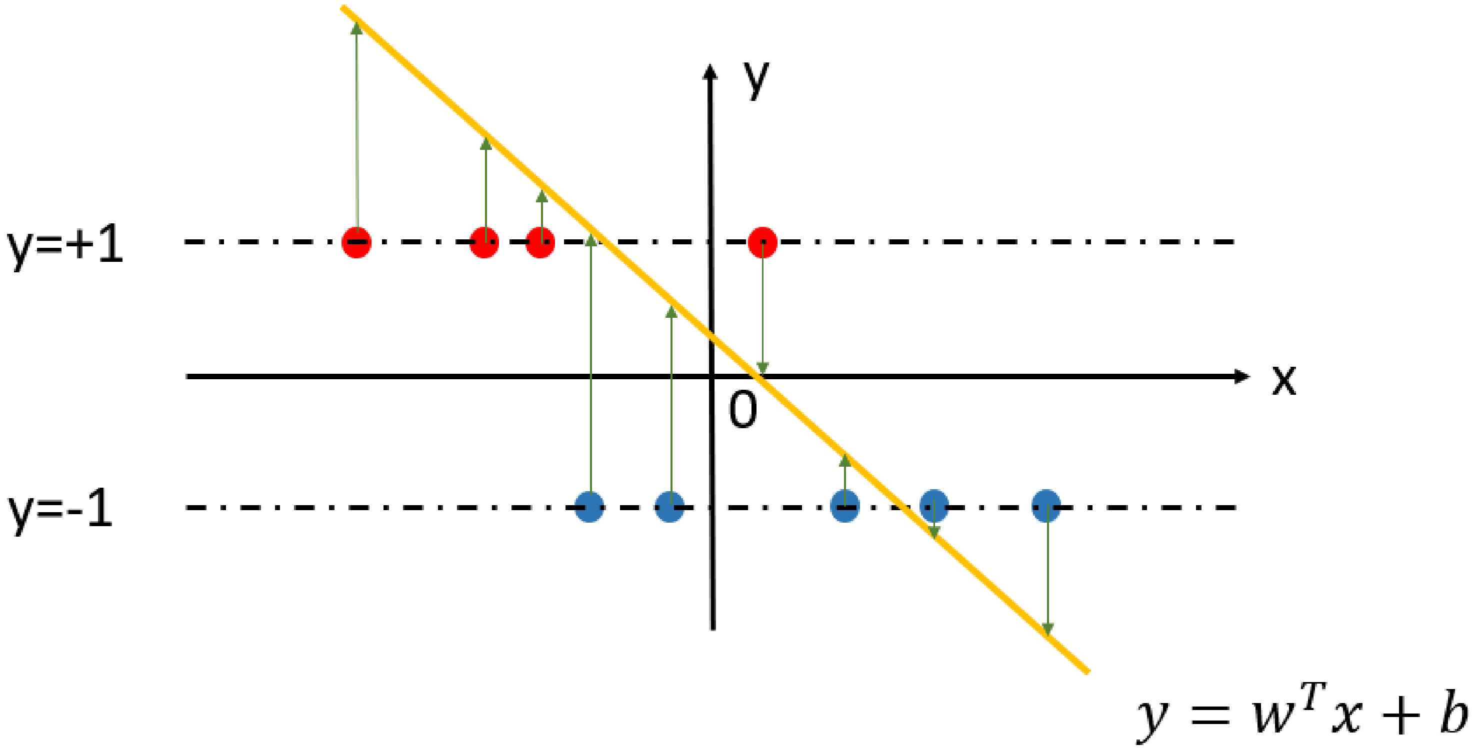
- By fitting a linear line of $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where

$$y_i = \begin{cases} +1, if\ the\ class\ is\ positive \\ -1, if\ the\ class\ is\ negative \end{cases}$$

  - We obtain the linear decision boundary of the input space

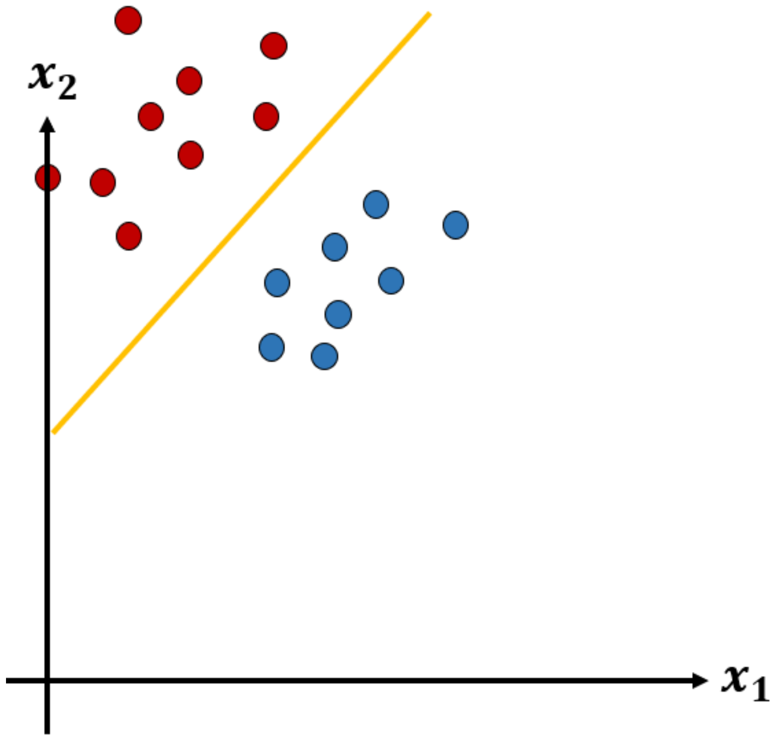
Fisher Classifier

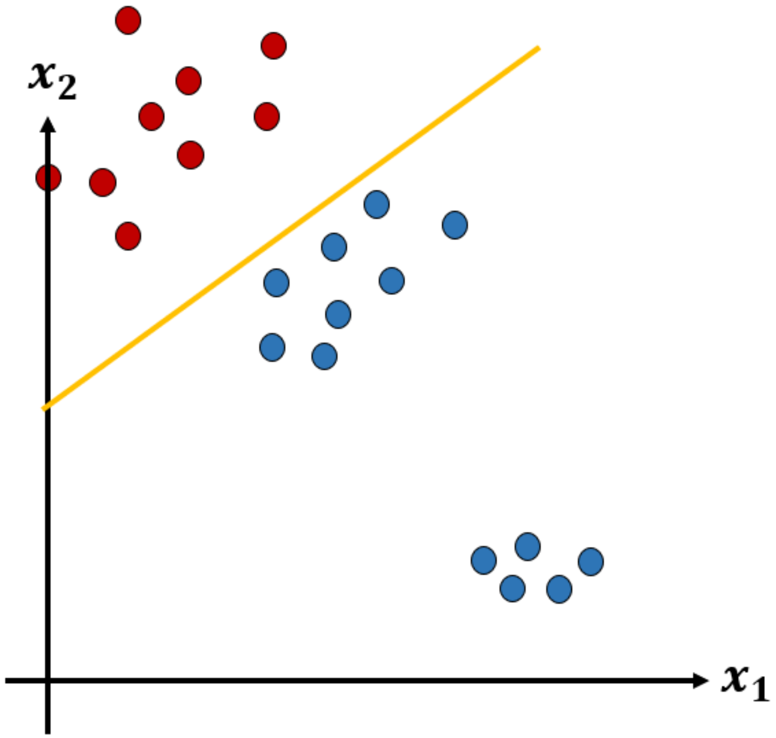
- Solution can also be given by least squares



Fisher Classifier

- Standard Linear Classifier (least squares) is sensitive to data distribution





Today's Agenda

- Previous Lecture: Supervised Learning

- Bayes Classification
  - Probability Basics
  - Bayes Classifier

- Linear Classification
  - Fisher Classifier
  - Logistic Classifier ☜

Logistic Classifier

- Also known as logistic regression, although it is a model for classification rather than regression

- It assumes the posterior probability of y is a logistic sigmoid of a linear function of $\boldsymbol{x}$

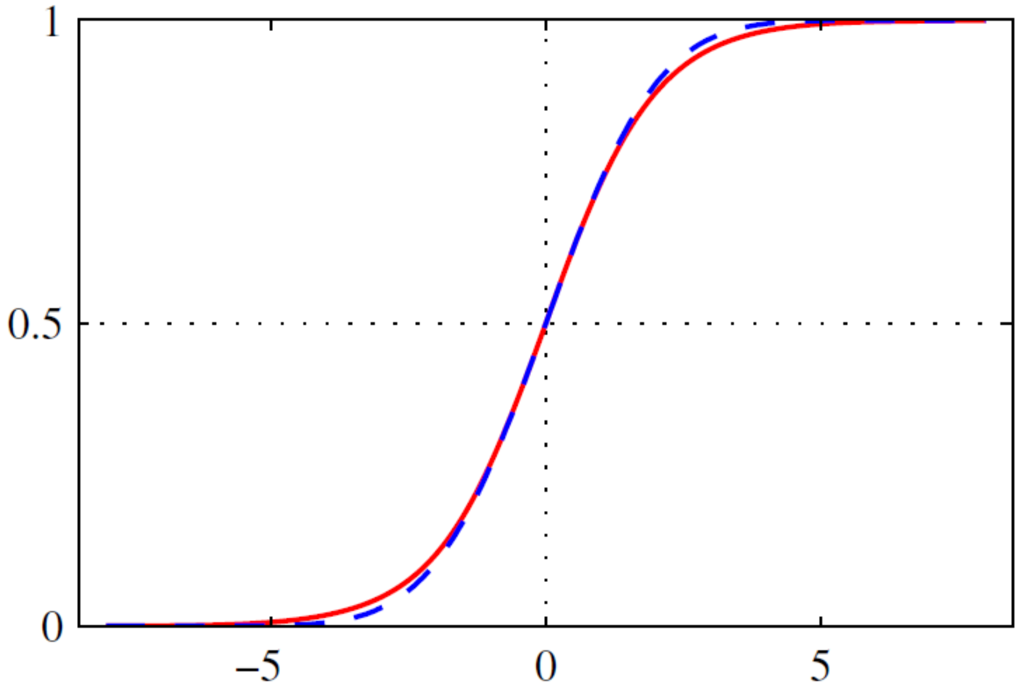
$$P(y|\boldsymbol{x}) = \sigma(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + b)$$

$$\sigma(f) = \frac{1}{e^{-f} + 1}$$

Logistic Classifier

$$P(y|\boldsymbol{x}) = \sigma(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + b)$$

$$\sigma(f) = \frac{1}{e^{-f} + 1}$$



Logistic Classifier

- Consider samples of $(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \dots, (\boldsymbol{x_n}, y_n)$ where

$$y_i = \begin{cases} +1, if\ the\ class\ is\ positive \\ -1, if\ the\ class\ is\ negative \end{cases}$$

  - We use maximum likelihood to estimate the parameters
  - The goal is to maximize:

$$P(\boldsymbol{y}|\boldsymbol{x}) = P(y_1|\boldsymbol{x_1})P(y_2|\boldsymbol{x_2}) \dots P(y_n|\boldsymbol{x_n})$$

Logistic Classifier

- Which equals to maximizing:

$$logP(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{n} logP(y_i|\boldsymbol{x_i})$$

  - If y_i=+1,

$$P(y_i|\boldsymbol{x_i}) = \frac{1}{e^{-f(\boldsymbol{x_i})} + 1}$$

  - If y_i=-1,

$$P(y_i|\boldsymbol{x_i}) = 1 - \frac{1}{e^{-f(\boldsymbol{x_i})} + 1} = \frac{1}{e^{f(\boldsymbol{x_i})} + 1}$$

Logistic Classifier

- Which equals to maximizing:

$$logP(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{n} log \frac{1}{e^{-y_i f(x_i)} + 1} = -\sum_{i=1}^{n} log(e^{-y_i f(x_i)}+1)$$

- Therefore, the problem transfers to minimizing

$$\sum_{i=1}^{n} log(e^{-y_i f(x_i)}+1)$$

Logistic Classifier

$$\sum_{i=1}^{n} log(e^{-y_i f(x_i)} + 1)$$

- Robust to outliers
- Can be solved by gradient descent
- There is no close form solution
- The solution really depends on the initialization etc.

Conclusions

- Many classification or regression problems can be specified as:
  - Find a suitable model / hypothesis
  - Define a loss function (i.e., least squares, maximum likelihood …)
  - Feed the data samples into the model and search for the model parameters that cause the least loss