Regression

* Continous Supervised Machine Learning.

* when we refer to Continuous here it refers to out put.
* but we can also have discrete out puts.

* Input is continuous

Continuous VS Discrete output

eg. Age : Continuous since time is continuous
* weather : Sunny or Rainy : discrete. Although there are other variables related to weather that are discrete outputs
In reality weather is continuous.
* In an argument most situations that are discreate are continuous.

* Phone number : Discrete

* Person wrote an email : Discrete
* Income how much you make : continuous is \$10K/mth ≈ \$9.9K/mth.

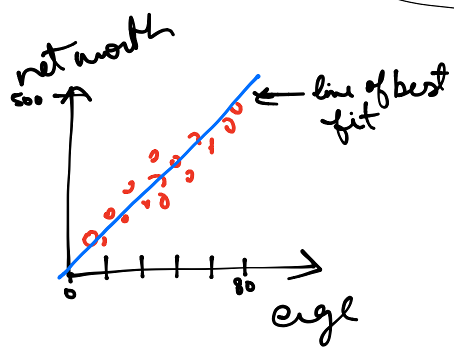
Ex: Terrain classification (Self driving Car)

* Continuous output → speed of the car

*

Discrete Out	Continuous Output
fast/slow	* Speed of Car in mph
	a good generization for

ex:



net worth ⇒ $y = mx + c$

N·W = 6.25 Age + C → intercept

N·W = 6.25 Age → Input

↓ slope

output → target variable

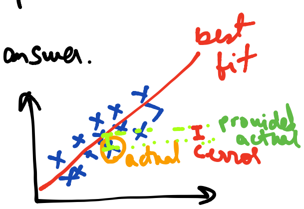
* $y = mx + C$

Intercept C, is the y value when $x = 0$.

* you can use ex: reg = LinearRegression().fit(X_train, Y_train)
to fit the training data.
* R^2: you can use reg.score(X_test, Y_test) to find out the coefficent of determin determination of the linear regression.
* reg.score(X_train, Y_train) to find out R^2 for train data set.

* Linear Regression Error:
eg: error is actual answer vs what is predicted

* error : actual answer – Predicted answer.

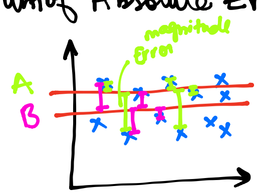


* what does a good fit minimize .
* Sum of Absolute error $\Sigma|\text{Errors}|$
* Sum of Squared errors $\Sigma|\text{Errors}|^2$

* In linear regression we use minimize sum squared errors and gradient descend to m minimize errors.

* Weakness of Sum of Absolute Errors:

predicted lines



* even though the sum of absolute errors is the same in both the predicted lines A & B.
* Line B represents a better fit, therefor use of Sum of Squared errors is a better choice to determine the line of the best fit
* There could be many lines that minimize $\Sigma|\text{error}|$, but only one line will minimize $\Sigma\, \text{error}^2$

* AS you add date SSE data becomes larger since more data point are taken into consideration, even though the fit might be good. Thee for SSE is not a effective evaluation metric.

* R^2: a effective evaluation metric
R^2: is how much change in my output (y) is explained by my change in input (x)

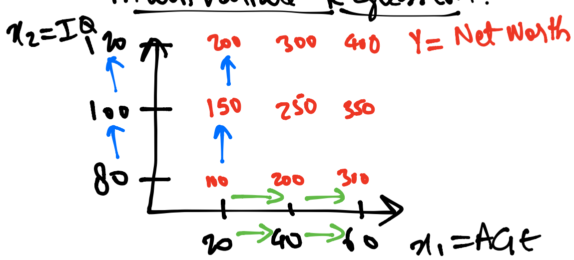
$0.0 \leq r^2 \leq 1.0$ ← doing a good job

↑ isn't doing a good job

Comparing Classfication & Regression

property	Supervised Classification	Regression
output type	discrete	Continuous (numbers)
what are you trying to find	decision boundary	"best" fit line
evaluation	accuracy	SSE "sum of squared errors" or R^2

Multivariate Regression:



$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

* x_1 : as x_1 increases by 20
  y increases by 100

* $x_2 \Rightarrow$ as x_2 increases by 20
  y increases by 50

* $m_1 x_1 \Rightarrow \frac{\cancel{100}\,5}{\cancel{20}} x_1 \Rightarrow 5x_1$

* $m_2 x_2 \Rightarrow \frac{5\cancel{0}}{2\cancel{0}} x_2 \Rightarrow \frac{5}{2} x_2$

$$y = 5x_1 + \frac{5}{2}x_2 + c$$

$$(20, 80, 100)$$

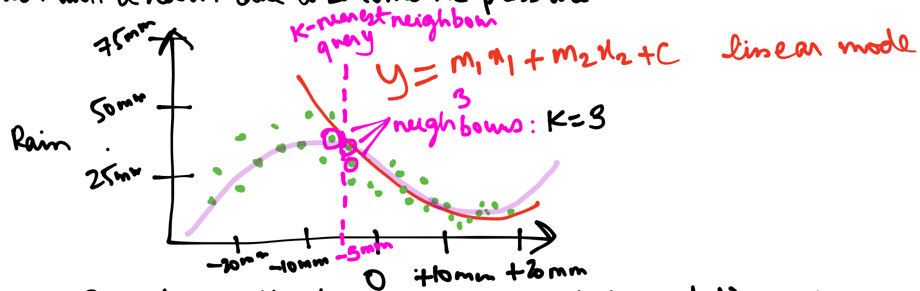
$$c = -200$$

$$y = 5x_1 + \frac{5}{2}x_2 - 200 \quad //$$

* Regression : can also be called Numerical Model

* Parametric Regression : we represent using parameters : $y = m \underset{}{\overset{\text{function}}{x}} + C$

* eg : barometric pressure → low : bad weather comming , high : good weather
* how much will it rain due to barometric pressure.



Instance Based methods: (Non Parametric model) (we toss the data after finding the values in the eq)

* K-nearest neighbour : a data centric approach, instance based approach where we keep the data when we make a "query".

* suppose we want to find out how much will it rain when barometric reading is -5mm. We will make a query at -5mm and find 3 nearest neighbours (K=3) to esti estimate how much it will rain at pressure -5mm.

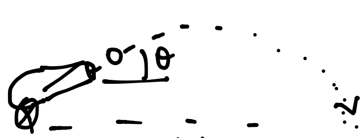
* How to Predict K-nearest neighbour : we take the mean of y values

* if we do it over alot of point we will get a nice "smooth hyperbola"

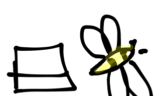
* Kernal regression: we weigh the contribution of each neighour point to how far they are in distance.

where as in K-nearest neighbour we take equal weightage of each neighbour point

Q Parametric (eq. based) or Non Paramatric mode) (data points based)



projectile ✓



will bees be attracted to the food source as its richness increases? ✓

y = how many bees?

x = food source richness

* Projetile is wel defined by a engineering equation, we can apply mathematical equaltion. Parametric

* Bee is not defined by a mathematical Solution. (Non-Parametric)