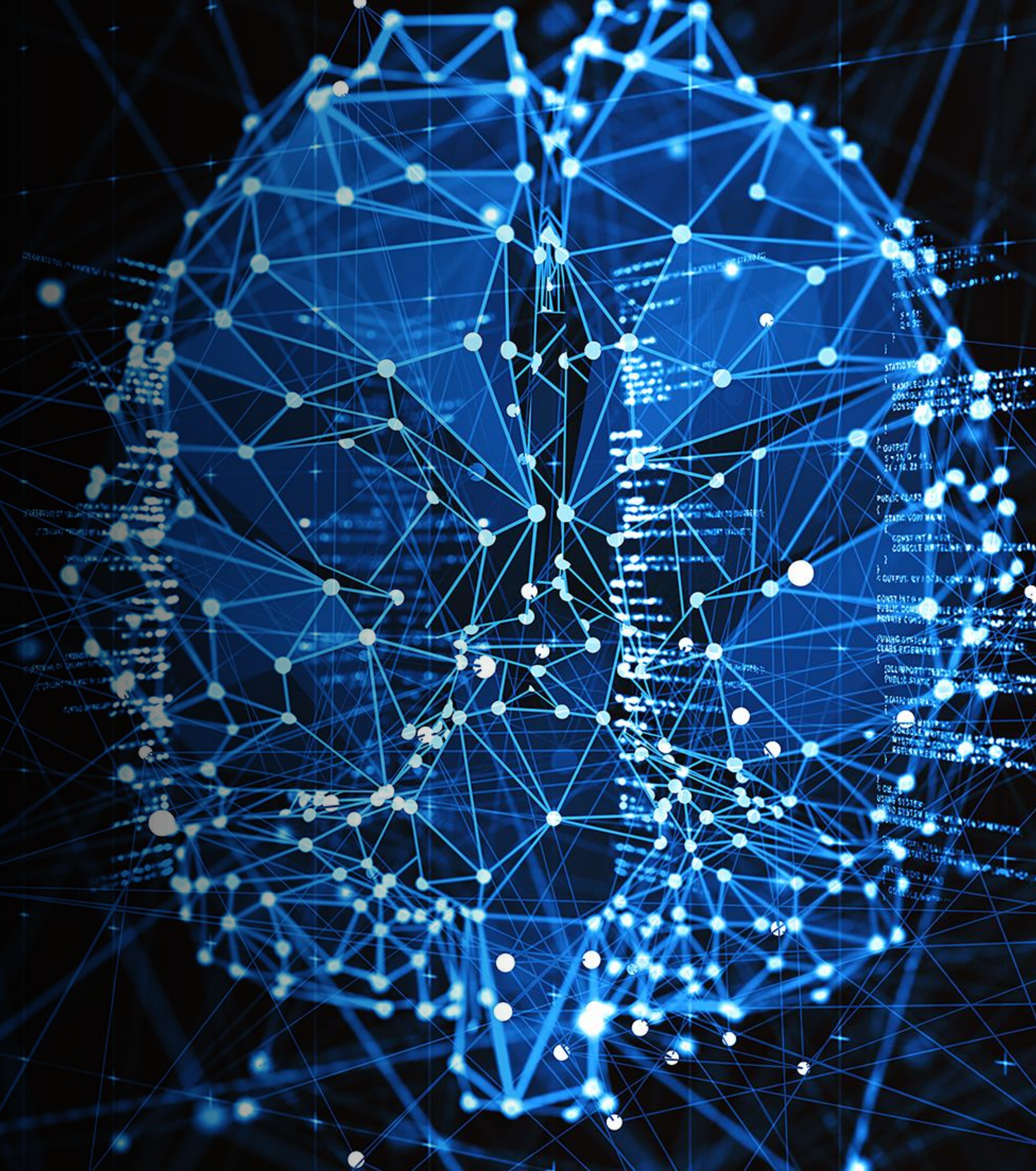
Thursday Learning Hour

Towards Reliable Machine Learning

Prabakaran Chandran

20-01-2022



Thursday Learning Hour - 20-01-2022

“ True Wisdom is knowing what you don’t know“

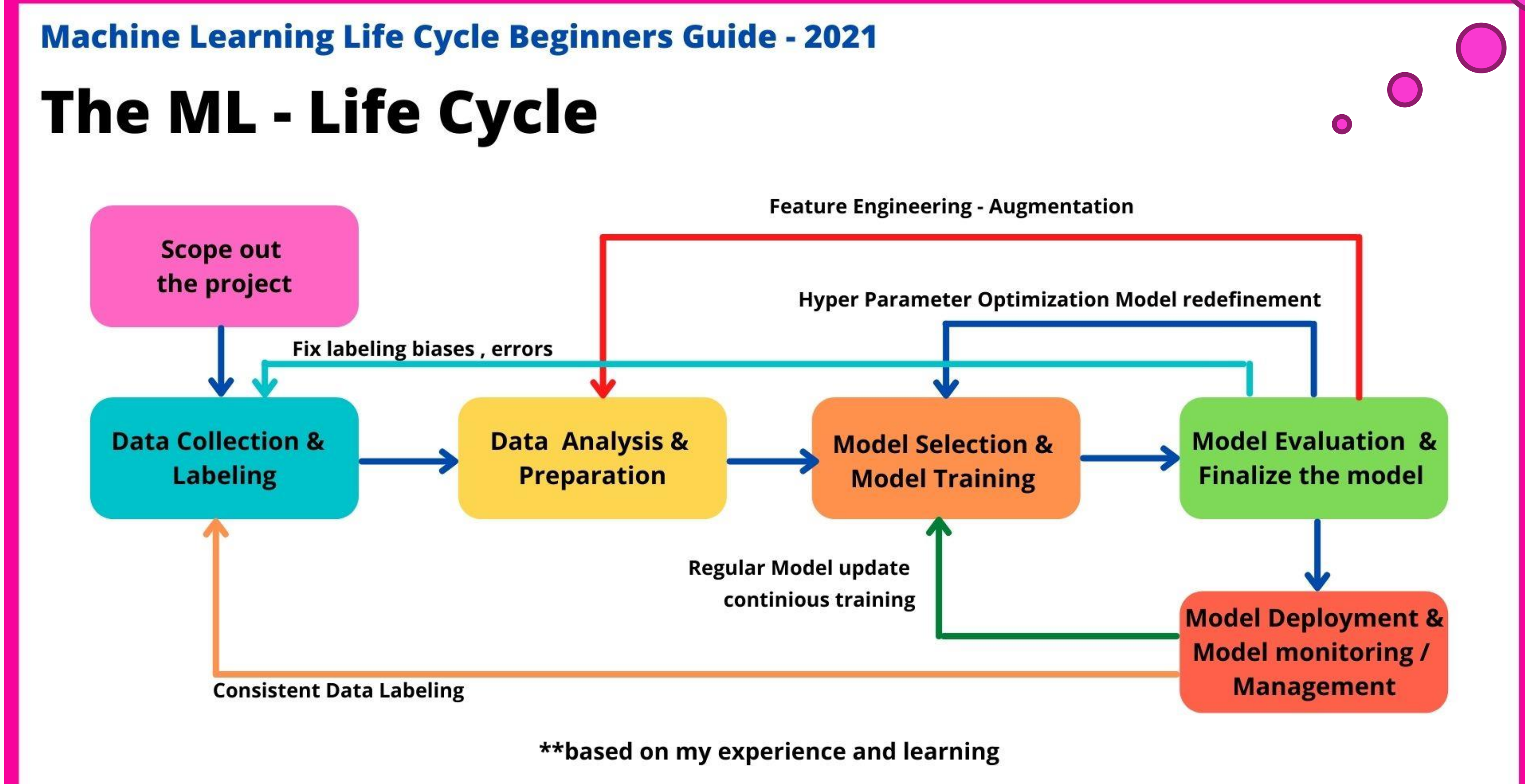
Confucius

What are we going to learn today!

- Contemporary Machine Learning Life cycle
- What is Reliable Machine learning
- Uncertainty matters!
- Quantile Regression
- Introduction to Conformal Prediction

Contemporary ML Life cycle

We are
missing
something?



This is from one of my LinkedIn Posts – Refer to that

What makes a Machine learning Exercise Reliable ?

Reliable Machine Learning

- Robustness
  - Continuous Training and Monitoring
  - Production in Scale
  - Withstanding Adversarial attacks
- Trustworthy/Quality
  - Good Performance Metrics
  - Fairness
  - Explainability (XAI became a trend post 2019)

** stability ,fairness, and explainability

We are much concerned about !

1. Good ML Ops Cycle – (This become a must to have option)
2. Well tuned Hyperparameters – (We are already doing and our last pitstop also)
3. Drift Analysis and Postproduction Monitoring (We TDS are not getting this chance all the time)
4. Scaling to the Broad usage -- (Depends on client's requirement)
5. Standard Infrastructure Needs – (Depends on Client's Capability)
6. Good Performance (Obviously) – There is a loophole!
7. Combating Adversarial Effects/Attacks

8. Some time Explainability (Post 2019)

Then…..

Then………..

…………………?

We are missing something out there!

Not having a More Concern about

Uncertainty of the Models / Quantifying the Uncertainties

Btw, What is that? Are We considering that as an Important thing?

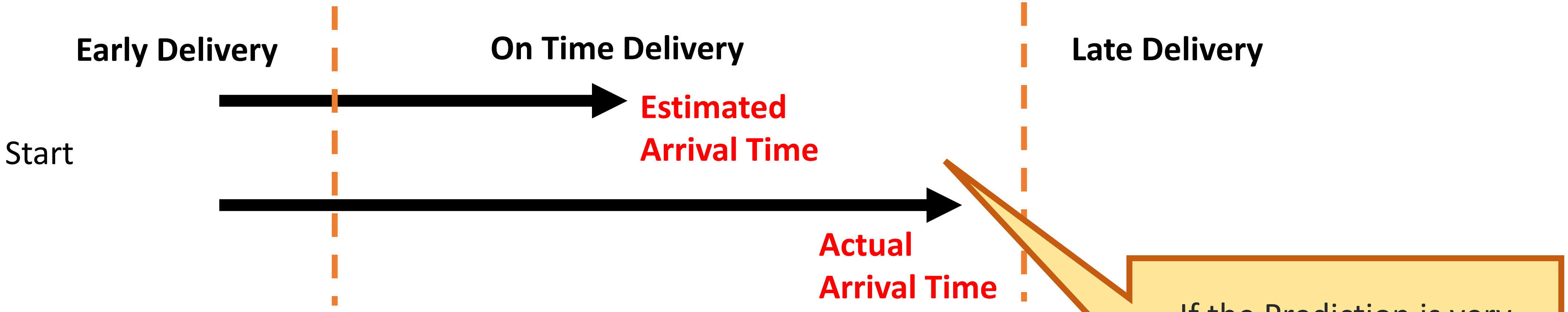
Let’s Embrace the uncertainty for better decision-making

Let’s Consider the Scenario – Price Prediction

- Problem : To predict the Price of an Asset / Commodity / Product
- Example : Real Estate Price Forecasting, Stock Price Prediction
- Situation : High stochastic Market in Nature (Especially Stock Market)
- What we do ?: We build General/sophisticated Algorithms (Imagine From Linear Regression to Deep Neural Transformers)
- What we predict ?:We estimate of Price values (a Point prediction with confidence
- What we miss? : Quality and Quantity in terms of Uncertainty

** add something

Let’s Consider the Scenario – Time Prediction



Actual Arrival Time = ETA + Prediction Error < Due Time

- We don’t Know the future prediction error.
- Stochastic Process can’t be handled easily
- Though our Model is not so biased , It is correct/Confidence on Average not on the error

What we need is “A buffer which can account Uncertainty?”

ETA + Buffer < Due Time

The buffer needs to be high enough to cover the risk of Uncertainty in most of the cases.

Why Uncertainty matters ?

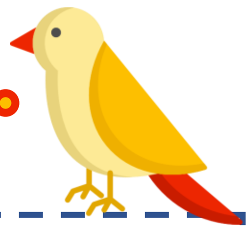
Decision Making Process should be supported with a Prediction with a level of Confidence

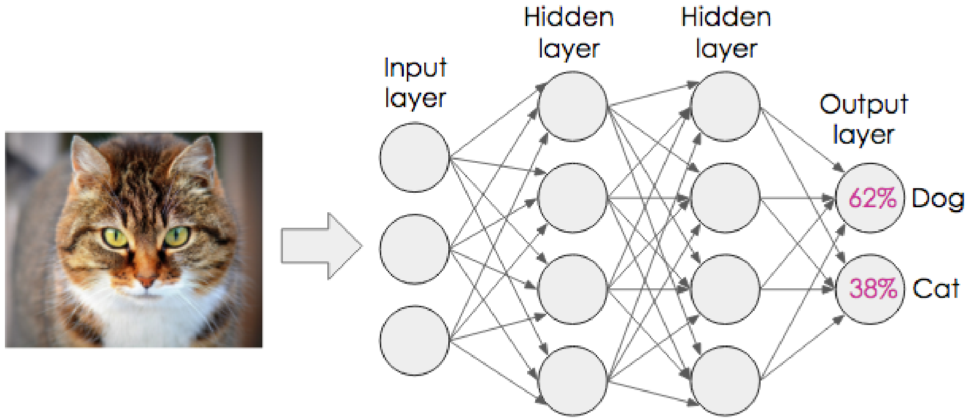
You may tell , We have Error metrics to measure accuracy .

But all of them are global level of confidence metrics.
So, it is necessary to look for

“Local Levels of Confidence at the Sample Level “

Hey There!
I am Chittu kuruvi!
(Lil Sparrow)
Will I be classified
as Cat/Dog





It should be 100% Cat :(

What If , the input image is a Bird ?

Will it be classified as “Dog” or “Cat” ?

YES, the Neural network still predicts as “Dog/Cat” with a high probability score on unseen images known as *“out-of-distribution”* samples.

What Causes Uncertainty ?

1- Approximation: since the model is not sufficiently expressive to model the data-to-label association.

2- Aleatoric: due to the intrinsic stochastic nature of the association. Aleatoric uncertainty captures noise inherent in the observations. This could be sensor noise or motion noise, *resulting in uncertainty that cannot be reduced even if more data were to be collected*

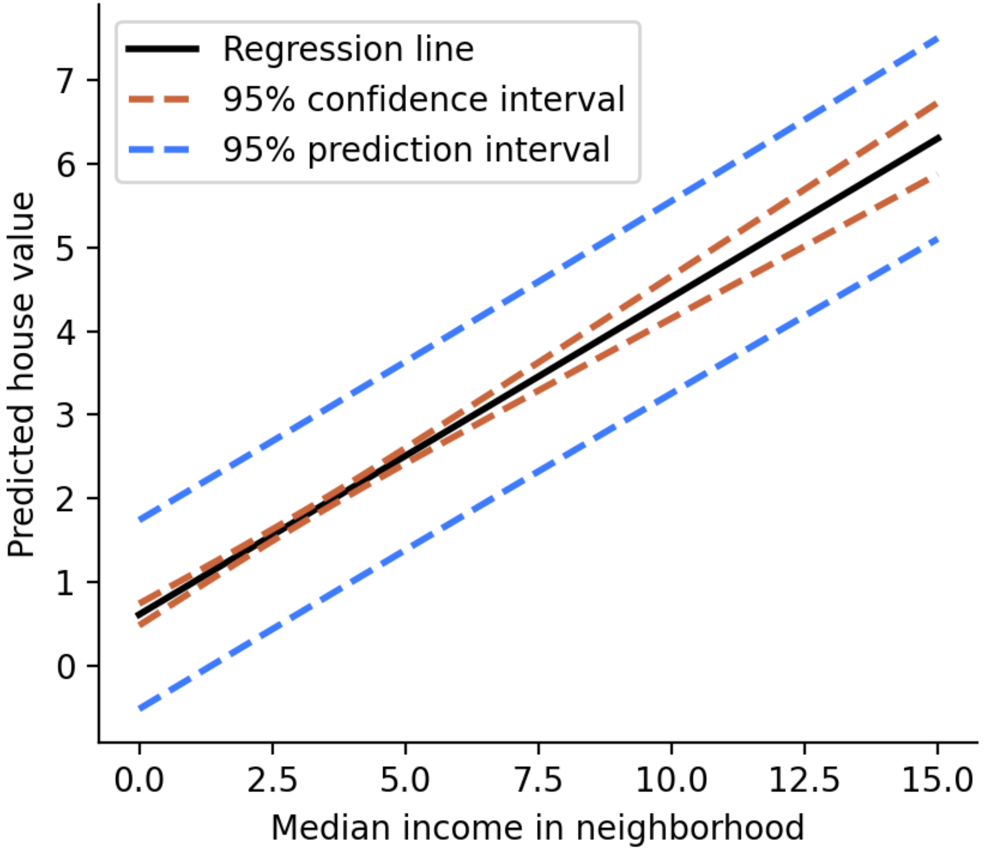
3- Epistemic: accounts for uncertainty in the model parameters — uncertainty, which captures our ignorance about which model generated our collected data. This uncertainty can be explained away given enough data and is often referred to as *model uncertainty*, *reduced given enough data.*

•measurement error in the data
•sampling error due to the inherent stochasticity of data collection process
•estimation error in the modeling process
•model misspecification error etc.

Basically, confidence intervals and prediction intervals quantify uncertainty in statistical estimates

***Simplify them*

What are the Few Ways to Quantify Uncertainty?



How certain are you that the point estimate is the actual value you are trying to predict?

Confidence Interval

Confidence Intervals are estimates that are calculated from sample data to determine ranges likely to contain the population parameter(mean, standard deviation)of interest.

Prediction Interval :

The range that likely contains the value of the dependent variable for a single new observation given specific values of the independent variables, is the prediction interval.

This is What we need ? But in a Reliable way

Quantile Regression - 1978

Linear Regression : A method of least squares to calculate the conditional ***mean*** of the target across different values of the features

quantile regression estimates the conditional ***median (50th Percentile – Q3 Quantile)*** of the target.

Quantile regression is an extension of linear regression that is used when the conditions of linear regression are not met (i.e., linearity, homoscedasticity, independence, or normality)

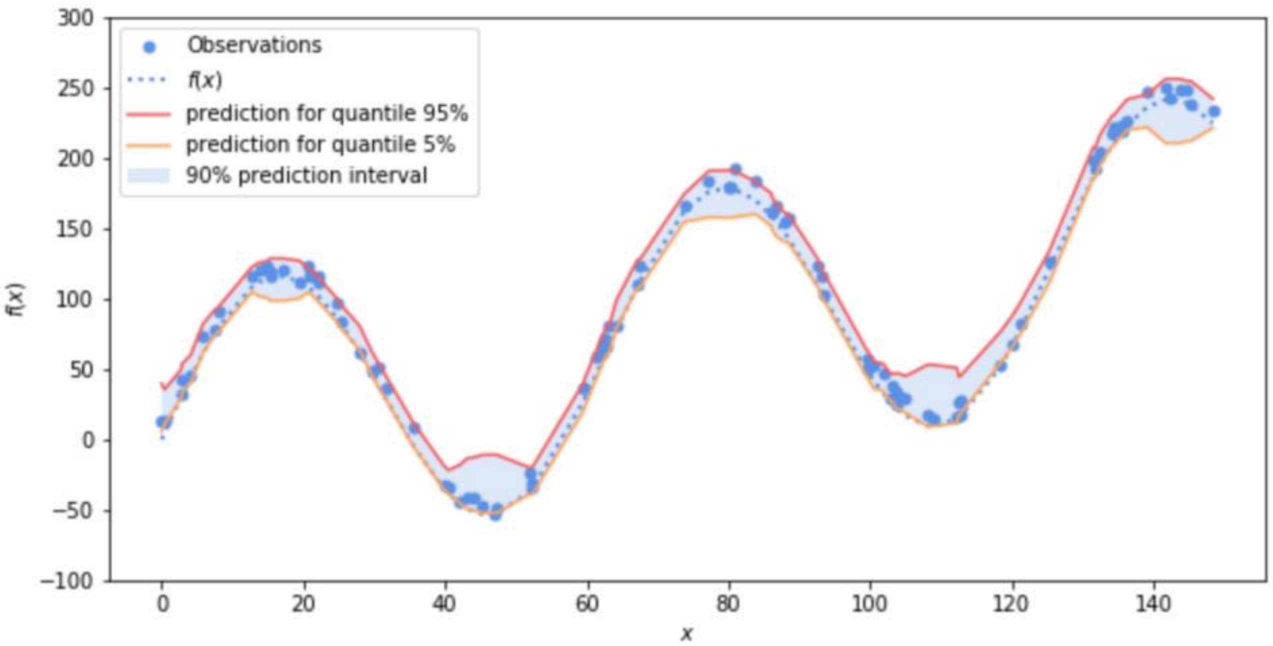
$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \quad i = 1, \ldots, n$ To Reduce the Error : $MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) \right)^2$

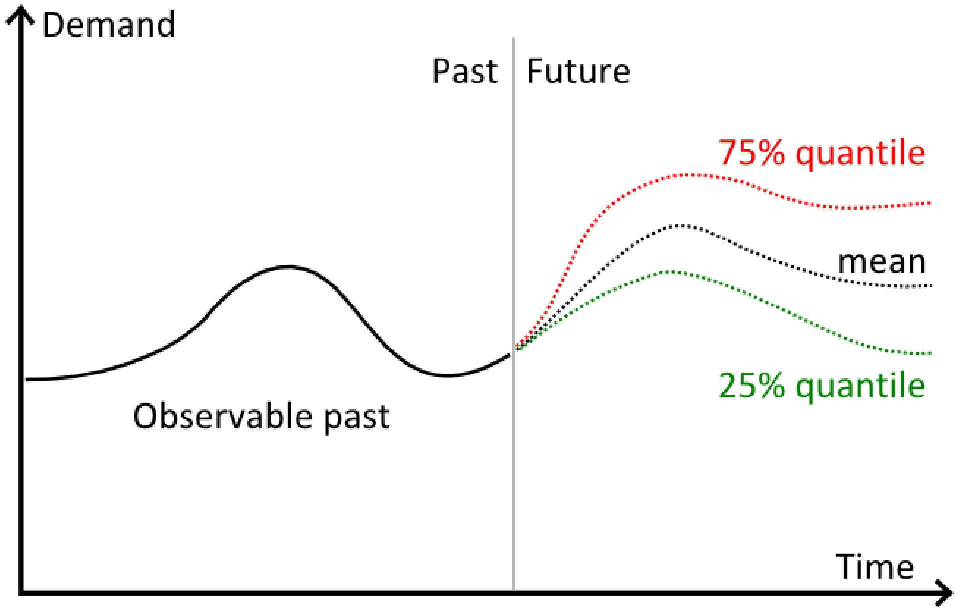
$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau) x_{i1} + \cdots + \beta_p(\tau) x_{ip}$$

Where p is the number of regressor variables
n is the number of data points

To Reduce Mean Absolute Deviation $MAD = \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(y_i - (\beta_0(\tau) + \beta_1 x_{i1}(\tau) + \cdots + \beta_p(\tau) x_{ip}) \right)$

Quantile Regression





From the Estimated Quantiles We can Evaluate the Best-Case and Worst-Case Scenarios

A quantile is the value below which a fraction of observations in a group falls. For example, a prediction for quantile 0.9 should over-predict 90% of the times.

This Big boy of 40+years old need some upgradation

Why do we need to Upgrade Quantile Regression ?

If you remember Chittu Kuruvi' s Scenario , Which is Out of Sample / Out of Distribution Case ,

Quantile Regression Can't Work in that case!

That's Where Conformal Prediction Come into Picture!

Conformal prediction only requires one assumption called exchangeability.

Exchangeability is the notion that any ordering of the data are equally likely to occur. (This became a skeptical thing also – Let's see that latter on!)

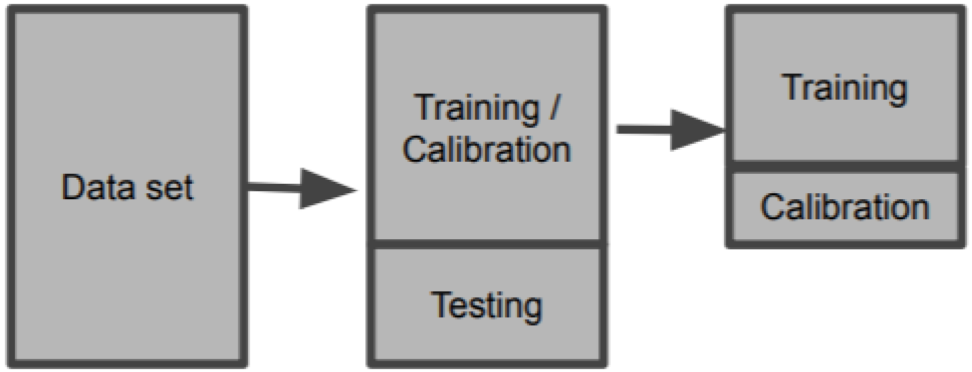
Conformal Prediction

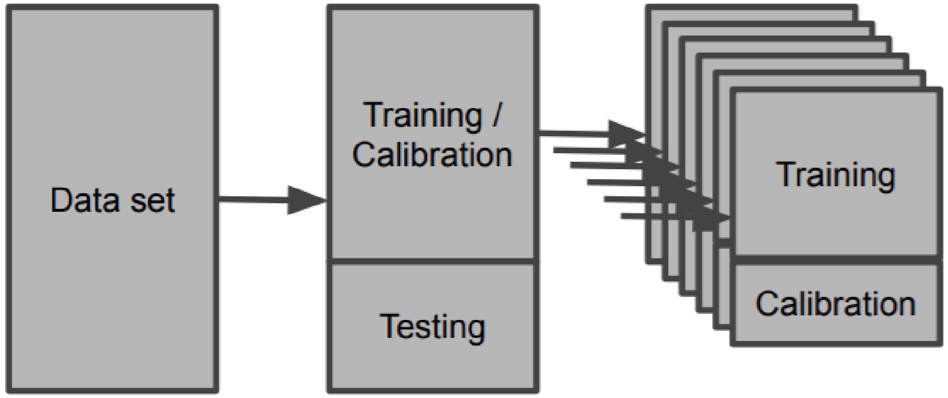
- It is a Model agnostic Way of Calculating Uncertainty
- Distribution Free , So work on any Out of Sample Things
- **Any model can incorporate conformal prediction** – Be it a Linear Regression Model or a Neural Network Model
- Just a Post hoc Calibration Exercise
- For Regression Problem : It Provides Prediction Intervals (Point Prediction to Set Prediction)
- For Classification Problem :   The single class prediction to a set prediction. If we have multiple classes in the Prediction set , It means model is under performing . But Usually , If the Point predicted Probability is 30% and Argmax, we still believe model is good (Based on Model's Global Performance Index)

Conformity / Non-Conformity

A non-conformity score measures how much each record doesn't conform with the rest of the data.

For a Regression Process , Non-Conformity/ Conformity Score = Prediction ~ True Calibration (Just Error)





Let’s Start : Calculate Conformal Scores

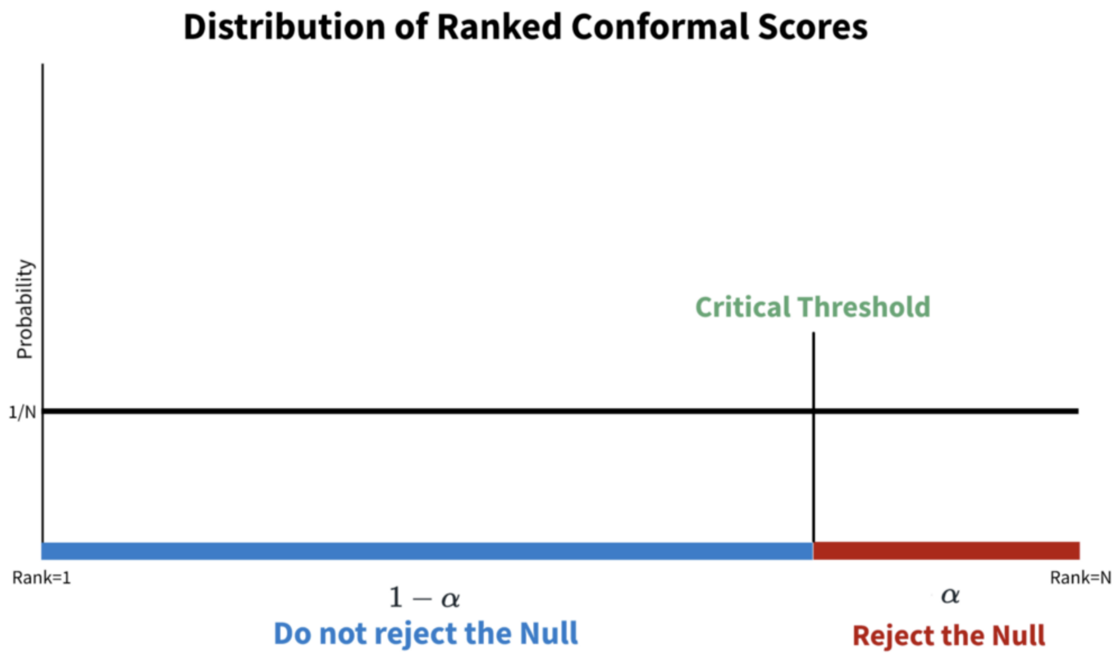
$$s_i = y_i - \hat{p}_i(y_i | X_i)$$

- s_i: conformal score
- y_i: binary label corresponding to **“ Cat or Dog “**
- $\hat{p}_i$: predicted probability by our model
- X_i: predictors
- $_i$: subscript corresponding to the index of our observed data

- After fitting the model, Calculate probabilistic predictions of our dependent variable . With the help of Predicted Probabilities
- Conformal Scores are calculated using the testing set, which is called the *calibration set*

Then , Get the Prediction Interval

- After getting conformal scores for all labels in our calibration set, order the absolute value of the conformal scores from low to high



- **Critical Threshold = 1 – alpha**
- **Alpha – Significance value**
- **If alpha is -.05 , Critical Threshold will be 0.95 ,**

- **Then the values inside the 95 % are significant and inside prediction interval , red values are outside, so they are outside the interval**

- **With the help of this significance , we can label each predictions with the Quantile measure**

Then , Estimate the Probability of Each Label

- Now We have Conformal Score and Critical Threshold
- If Probability is less than Critical Threshold, It is True
- If the Probability is greater than Critical Threshold, It is False
- Simple Binary Classification Models (Logistic Reg or Neural Nets are mostly on Argmax (SoftMax classifiers – give Label (class) to Class with High Probability
- But with conformal prediction, we can allow the model to say that neither or both are the true label. In effect, the model can say “I don’t know.”
- We are not Forcing our model to predict with an outcome , we get much more robust forecasts in our prediction sets.
- So Chittu Kuruvi will be predicted neither of the options

Then , Note Some thing

- This is a very earlier way of Conformal Prediction
- Academia have developed lot of Conformal Predictors especially in 2020 – That went exponentially
- This is expected to behave same in the Industry also in this year

- Few Interesting Python Packages where you can try this : Orange , Nonconformist

- Follow Royal Holloway / Carnegie Mellon Profs/Fellow to learn more

Questions ?