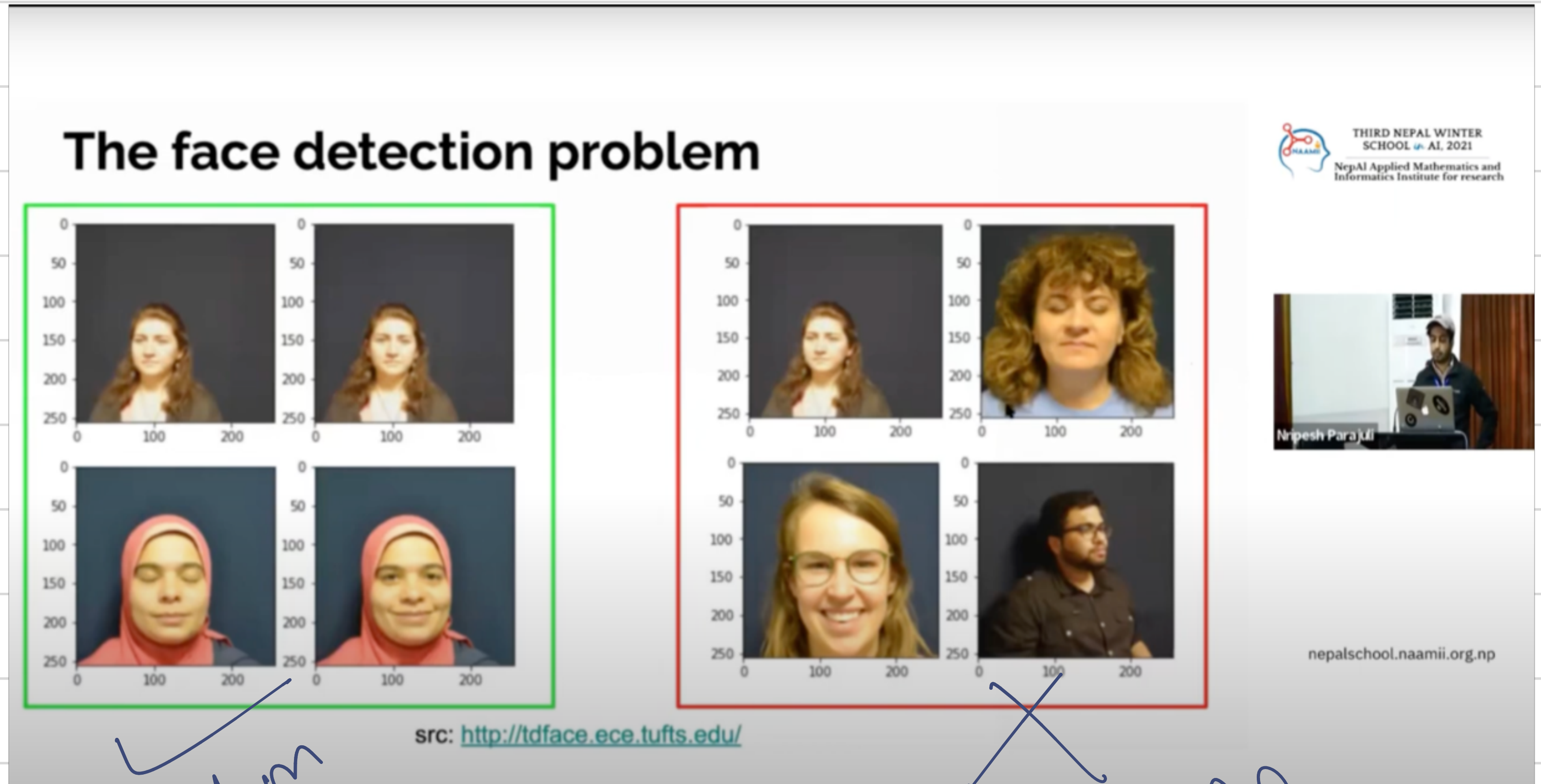
Day 18, Oct-20, 2024

Face Detection using CNNs.

The first step to face detection is classifier with bounding rectangle box.

then only we move further. For an example, facebook tagging, Camera/Smart phone capturing face.



Detection Correct

Detection Incorrect

Facebook does face classification because it has billion of users and billion of data. But, does Facebook have billion parameter model just to detect the face of every person? What if, new person/user join the Facebook? Will Facebook build again new model?

So, How does Facebook Handle such a mass user face detection?

→ Answer is Featurization, geometry co-ordinates of a face projection, we need to make something (geometrical co-ordinates) like features.

And these features should be such that faces that are similar and in the feature space the distances is low and the faces that are far apart the distance is far and then you can do Clustering.

now if you have a well-defined distance metric on this space let's call this a face space. We cannot exactly say what does the facebook face detection work. But, they must really optimized model that can handle billion of users tagging and a face detection.

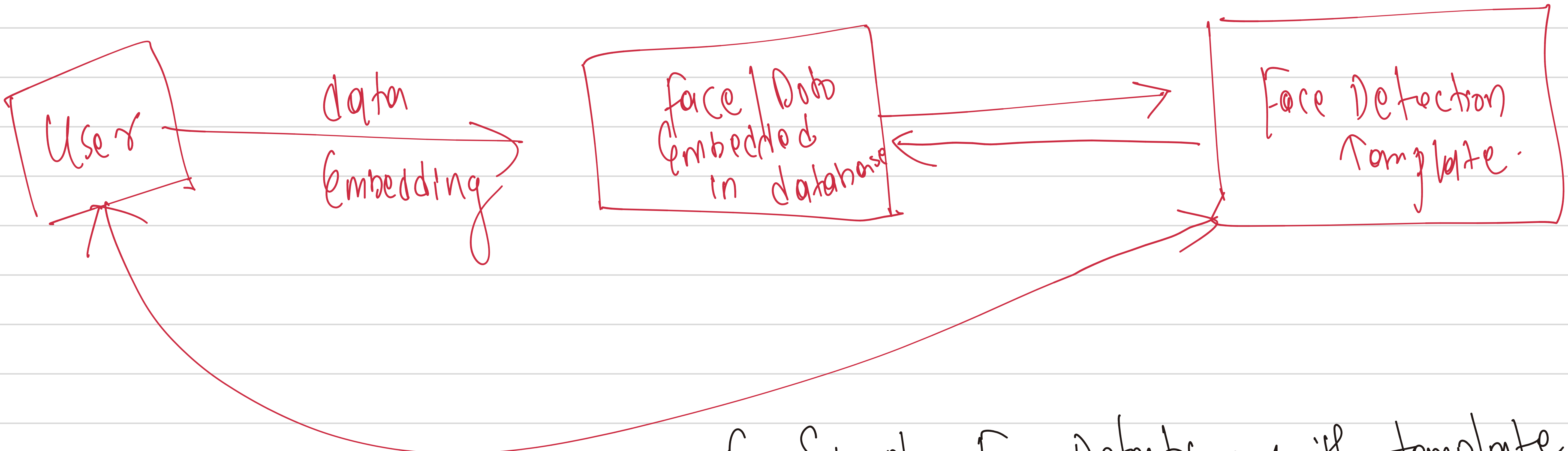


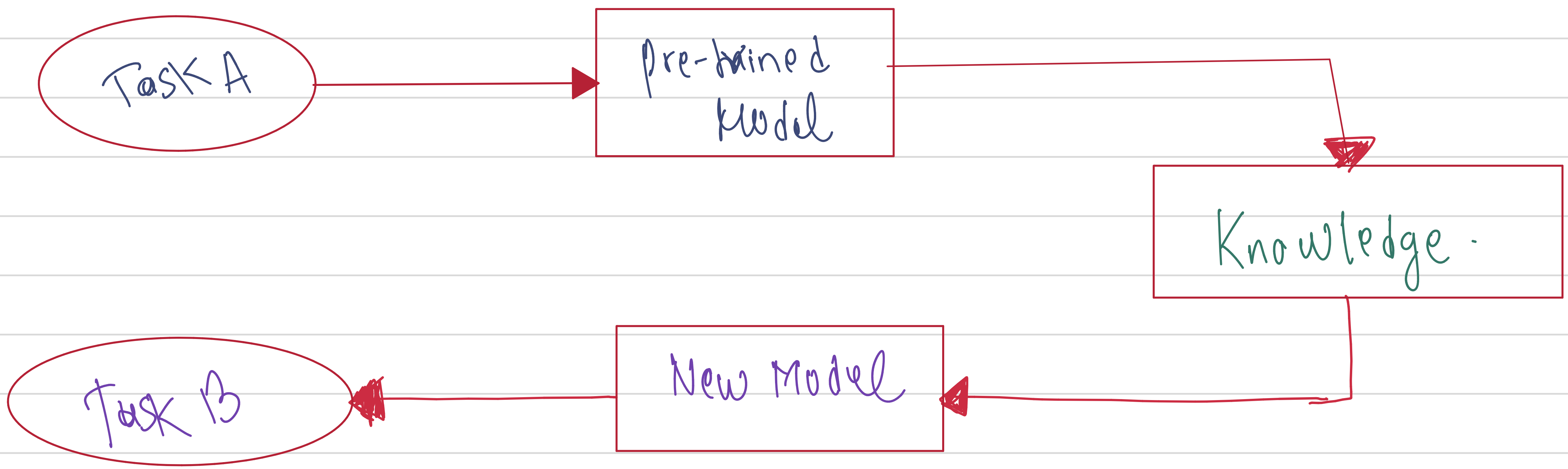
fig. Simple Face Detection with template
for large Scale.

So, the main thing is to remember that it is not good practice to build the face classifier because as correctly pointed out every a new person comes it is pointless so what you want to try and do actually is build this thing called an embedding or just learn the feature.

Transfer learning:

TL, using only classifier?

Transfer learning is an AI technique where a model parameter trained on one task is adapted to perform a different, but related task.



Instead of starting from scratch, the model leverages knowledge from the original task, improving efficiency and reducing the need for extensive data.

Max pooling ignore the less important features so Max pool is translation variance.

Hit & trial-

Facebook does training like Classification for thousand people or uses ImageNet you still want this idea of distance because you cannot train a Classifier with billion of parameters. But your featured model can be trained model like VGG or Resonance or whatever model trained on Imagenet the final top Classifier layer you can take that and that could be your feature.

But what would be the problem with that?

We need to tune that to make work with the various faces.

Classifier for thousand Images → Fine-Tune the Model → Last top layer that could be embedding → New face

The Face Detection Problem

- Need to able to match faces across:
  - Pose
  - Facial Expression
  - Lighting Conditions
  - Background
  - Clothing items - hats, sunglasses, etc
  - And much more -

[Try to match | unmatch faces across pose + facial expression]

Classification vs Regression?

- While the face detection problem can be posed as a classification problem as well, another way is to pose it as a regression problem.

- Why Regression?

  - if we learn a regressor that computes the "distance" between 2 faces, this regressor is generalizable to new unseen people's faces.

- A classifier is only useful if we want to predict a new face but of someone already in the training example.?

So, classification is like absurd or what it is?

Summary:

① Model will have 2 input / faces compute the distance between 2 input faces and the distance is not just some basic but the learned distance that should be represented / representive and it should be discriminative. Discriminative is for the features.

Regression function:

$$E_W(X_1, X_2) = \| G_W(X_1) - G_W(X_2) \| \quad \text{— eqn } ①$$

for matching pairs of images, we want to push E_W to 0. For non-matching we want to push E_W to 1.

The regression function

THIRD NEPAL WINTER SCHOOL in AI, 2021
NepAl Applied Mathematics and Informatics Institute for research

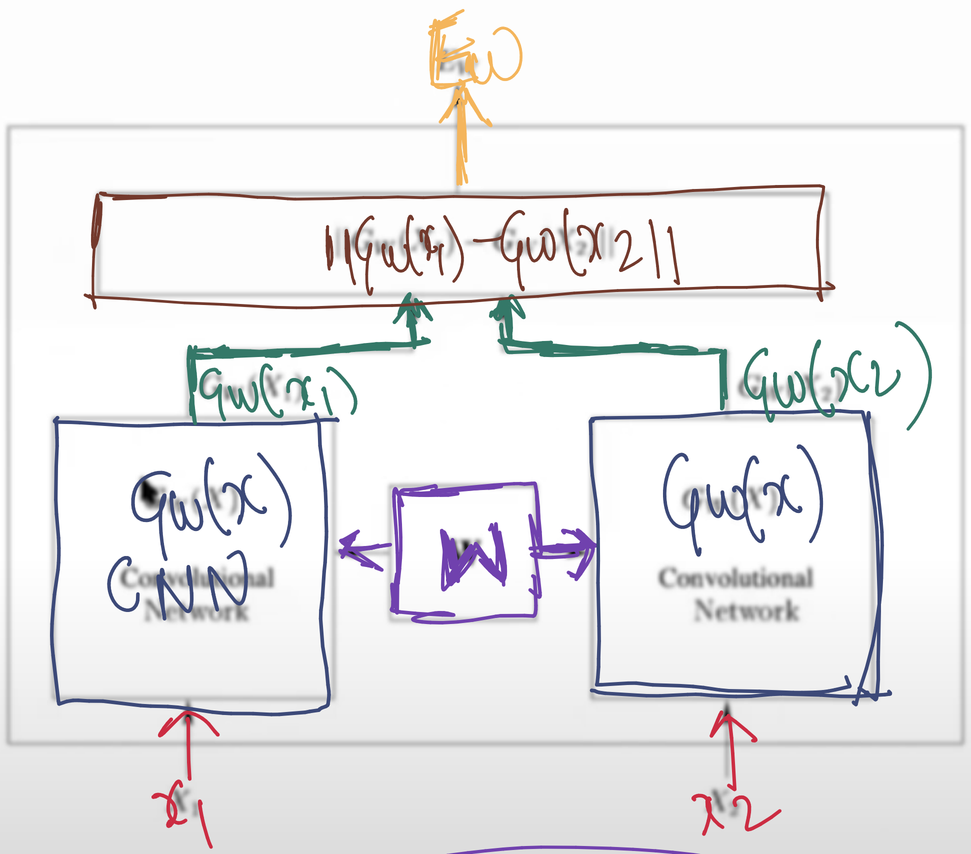


Figure 1. Siamese Architecture.

$$E_W(X_1, X_2) = \|G_W(X_1) - G_W(X_2)\| \qquad (1)$$

For matching pairs of images, we want to push E_W to 0. And For non-matching, we want to push E_W to 1.

Chopra, Sumit, Raia Hadsell, and Yann LeCun. "Learning a similarity metric discriminatively, with application to face verification." 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Vol. 1. IEEE, 2005.

nepalschool.naamii.org.np

1 x_1, x_2 → Input Image → 2 $G_W(x), G_W(x) \to W$ → 3 $G_W(x_1), G_W(x_2)$ → $\|G_W(x_1) - G_W(x_2)\|$ 4 → $\boxed{E_W}$ 5

Called as Siamese (twins). Architecture

Okay, Scenario where 'ω' can be Similar and Different?

Normal Images
- RGB color Images taken by Camera, phones

Other Types of Images
- MRI, CT Scan Images
- Infrared, Satellite Images
- Medical Imagings.

So, the W for the $q_w(x)$ can be Similar and Various ω according to the Input Images Nature.

From TextBook. (Propositional Logic: Knowledge Representation)

① Tautology:
Tautology is a statement that is true in every possible interpretation.

$p \vee \neg p$ (either p is True or p is False).

Tautology is Always true regardless of the truth values of the Components.

Either it will rain or it will not rain.
if it is Sunny, then it is Sunny.

$p \vee \neg p \rightarrow$ classical tautology.

② Validity:

Validity refers to the structural Correctness of an argument.

→ Depends upon the logical form, premises Must be true if only Conclusion is also true

→ Conclusion logically follows the premises.

Premise 1: All Humans are mortal

Premise 2: Socrates is human

Conclusion: Socrates is mortal

$\forall_x \; M(x) \longrightarrow S(x)$ So, $M(x) \rightarrow x$ is all Humans

$S(x) \Rightarrow$ Socrate is Human.

③ Well-Informed Formula (WFF):

→ WFF is a syntactically correct expression in formal logic.

→ It ensures logical expressions evaluated without ambiguity

→ checks syntax, rules, grammar in a statements / sentences.

Eg 1: $P \wedge Q \rightarrow R$

Eg: $\forall x \, (P(x) \rightarrow Q(x))$

→ Implies.

$\wedge \rightarrow$ and

$\vee \rightarrow$ OR

$\neg \rightarrow$ Not

Eg 3: $(P \rightarrow (Q \wedge R)) \vee (\neg P \wedge Q)$

P is peter of mine if he has two friends or do not have two friends.

4. Inference using Resolution:

Inference is the process of drawing conclusions from premises

(i) Resolution Method: It allows us to derive new clauses from existing clauses, mostly used in PL and First-Order Logic.

→ Deduce New clauses by eliminating a variable that appears in both clauses.

$(A \vee B) \wedge \neg B \longrightarrow A$.

- $A \vee B$
- $\neg B \vee C$

$\rightarrow A \vee C$ $\quad (A \vee B)\ \neg B \vee C \longrightarrow A \vee C$ ← New clause.

② Chaining → Forward chaining begins with known facts and applies inference rules to derive new facts until it reaches a conclusion or goal.

Process: start with facts "It is cloudy"

Apply first Rule to infer "It might Rain"

Apply 2nd " " "If it rains, the grass will be wet".

using Forward Chaining we can infer that grass will be wet

Backword Chaining:

→ Goal-Driven used in Logic Programming and works behind / backward, trying to find the premises that Support the goal,

Goal: Prove "the grass is wet"

Rule: "If it rains, the grass is wet"

Known fact: "It is raining"

Process: ① Start with Goal

② Work backward: check if it is raining

⑥ Since it is roining the goal is confirmed.

Forward Chaining:

Premise 1: P

Rule 1: $P \rightarrow Q$

Rule 2: $Q \rightarrow R$.

Use Rule 1 to infer 'Q' from 'P', then use Rule 2 to Infer 'R'

Backward Chaining:

Goal prove Q.

Rule 1: $P \rightarrow Q$

Rule 2: $R \rightarrow P$

If we know R is true, we can use Rule 2 to infer P, then use Rule 1 to infer Q.