NAME- Adarsh Salve
PRN - 22610076.

ANN Assignment - 4

Q1] Explain gradient descent algorithm in detail.

→ (1) Gradient descent is an optimisation approach used to minimise the loss function by iteratively updating the weight in direction of -ve gradient. Common variants include :-

i) Batch gradient Descent: update weight after completing gradient over entire dataset.

ii) Stochastic gradient descent: update weight after each training.

iii) Mini batch gradient descent: Update weight after completing gradient over small batch of training.

How it works :

for single row, we write loss function as :

$$J(w,b) = \frac{1}{n}(y_0 - y)^2$$

Here, x, y = input, w = weight, b = bias.

we differentiate partially wrt wt. & b.

$$J'_w = \frac{\partial J(w,b)}{\partial w} = \frac{\partial}{\partial w}\left[\frac{1}{n}(y_0 - y)^2\right]$$

$$= \frac{2(y_0 - y)}{n} \frac{\partial}{\partial w}[y_0 - y]$$

$$= \frac{2(y_p - y)}{n} \frac{\partial}{\partial w}[(xw^T + b) - y]$$

$$= \frac{2(y_p - y)}{n}(x - 0))$$

$$= \frac{1}{n}(y_p - y_p)[2x].$$

i.e. $J'(w) = \frac{\partial J(w,b)}{\partial w}$

$= J(w,b)[2x]$

Also $J'_b = \frac{\partial J(w,b)}{\partial b}$

$= J(w,b)[2]$

above is for linear regression and for fully connected network

param = param $- \gamma \nabla J$

γ = learning rate

J = loss fn

∇ = derivation of loss fn

param = wt and bias (multiple both)

steps:

1) find the gradient using loss.backward()
2) Get the parameter using model.linear.wt & model.linear.bias.
3) Update the parameter using formula.
4) again assign the model parameter to model.

Q.2) what is regulqnization in neural network? Give its significance

→ Regulanization is atechnique used in ML to prevent overfitting. overfitting happen ~~we~~ when a model learns the training data too well, including the noise and outlinears which causes it to perform poorly on new data. It adds penalty to the model for being too complex, encouraging it to stay simpler and more general. This way, it's less likely to make extreme prediction based on the ~~po~~

noise in data.

commonly used technique.

1) lasso Regularization (L_1 Regularization)
2) Ridge Regularization (L_2 Regularization)
3) Elastic Net Regularization (L_1 and L_2 combined)

significance:

1) Improve Generalization - Help the model perform well on unseen data.
2) control complexity - penalizes large wt to simplify the model.
3) Reduce overfitting - prevents the model from memorizing training data.
4) Enhance stability - Technique like dropout improve training robustness.

Q.3) What are different regularization technique used in network?

→

1) Lasso Regression (least Absolute shrinkage and selection operator). It adds the absolute value of magnitude" of coefficient as a penalty term to loss fn. It also helps in achieve featur selection by penalizing the wt to approximately equal zero if that feature does not serve any purpose in model

$$\text{cost} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{m} |w_i|$$

m = no. of feature

n = no. of example

y_i = Actual Target value.

$\hat{y}_i$ = predicted target value

2) Ridge Regression (L2)

It adds the squared magitude of the coe. as a penalty term to loss fn.

$$\text{cost} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{m} w_i^2$$

3) elastic net Regression

It implies that we add the absolute norm of the wt as well as squared measure of wt. with the help of hyperparameter that control ratio of L_1 and L_2

$$\text{cost} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda(1-\alpha)\sum_{i=1}^{m}|w_i| + \alpha\sum_{i=1}^{m} w_i^2)$$

Q.4) List the application of deep learning.

→ 1) computer vision
- object detection and recognition
- image classification
- image segmentation

2) Natural language processing
- Automatic text generation
- language traslation
- sentiment analysis.
- speech recognition.

3) Reinforcement learning
- Game playing
- Robotics
- control system

Q 5) Dissuss any one application in deep learning in detail

→

Medical image analysis - Brain Tumor Detection

How it works.

1) Data collection
- MRI images of brain are collected, labeled as normal or tumor.

2) Preprocessing
- image are resized, normalized and sometimes augmented (rotated, flipped) to improve model robustness.

3) model architexture.
- The CNN is trained on these image to learn pattern and feature like shape texture

4) Training and evaluation
- model learn to classify or segment tumor areas.
- performance is evaluted, precision, recall and F1-score.

5) output
- Trained output will detect whether a tumor is present and possibly highlight its location in image.

Q.6) List the varients of gradient desent learning.

→ 1) Batch Gradient Desent
2) stochastic Gradient Descent
3) mini-Batch Gradient Descent
4) momentum Gradient Descent

5) Nesterov Accelerated gradient (NAG).
6) Adagrad
7) RMSprop.
8) Adam (Adaptive Moment Estimation)
9) Adadelta.
10) ~~NaA~~ Nadam

Q.7) what is the need of convolution neural network in Deep learning

→ 1) Efficient feature extraction
CNN automatically detect features like edge texture, shapes and object from image using filter (kernel). the reduce the need for manual feature engineering

2) parameter sharing
Instead of connecting every neuron to every pixel CNN use the same filter across the image this reduces no. of parameters.

3) preserving spatial information
CNN maintain the spatial hierarchy in image.

4) Translation inversion
CNN are good at recognizing obj even if they appear in different position.

Q.8) Elaborate on CNN with advantange and disadvange of CNN.

→ Advantage.
1) Automatic feature extraction
learn feature directly from data without manual effort.

2) parameter efficiency
Used shares wt which reduce no. of parameter and speed up computation.
3) Better Generalization:
Pooling layer up help in making the model robust to translation.

Disadvantage:
1) Required large dataset
need large amount of data to train effectively.
2) computationaly intensive.
Training CNN demand high processing

Q 9) what is need of Recurrent neural network in deep learning
→ RNN are specially designed to handle sequential or time-series data. Unlike traditional network, RNN have a memory that capture information about previous i/p, making them ideal for task where order and context matter.

They are needed for.
1) Handling sequencial data.
2) Temporal dependency.
3) variable i/p & o/p length.
4) context awareness.

Q 10) RNN advantage, disadvange and application.
→ Advantage
1) sequencial information
2) Parameter sharing.

3) Flexible i/p o/p length
4) Real time processing.

Disadvantage
1) vanishing / exploding gradient
2) training difficulty
3) short term memory.
4) limited capacity

Application
1) text Recognition
2) sentiment analysis.
3) speech Recognition
4) music composition

Q.11) what are the drawback of RNN and how they are overcome?

→ Drawback
1) vanishing Gradient problem
2) exploding Gradient problem
3) limited memory
4) slow training
5) difficulty in handling long term context.

How to overcome it:
1) LSTM (long term short memory)
2) GRU (Gated Recurrent unit)
3) Gradient Clipping
4) Bidirectional RNN
5) Attention mechanism.

Q.12) write a note on long short term memory.

→ long short term memory is a type of RNN designed to overcome limitation of standard RNN

① memory : LSTM have memory cell that can maintain information over long period.

② Gate : forget, i/p, o/p

③ Better Gradient Flow

Application.

1) text ~~Recog~~ Generation

2) speech Recognition

3) machine translation

4) stock price prediction

5) music generation.