CIS 419/519 Recitation

Chang & Varun

Nov 24 2020

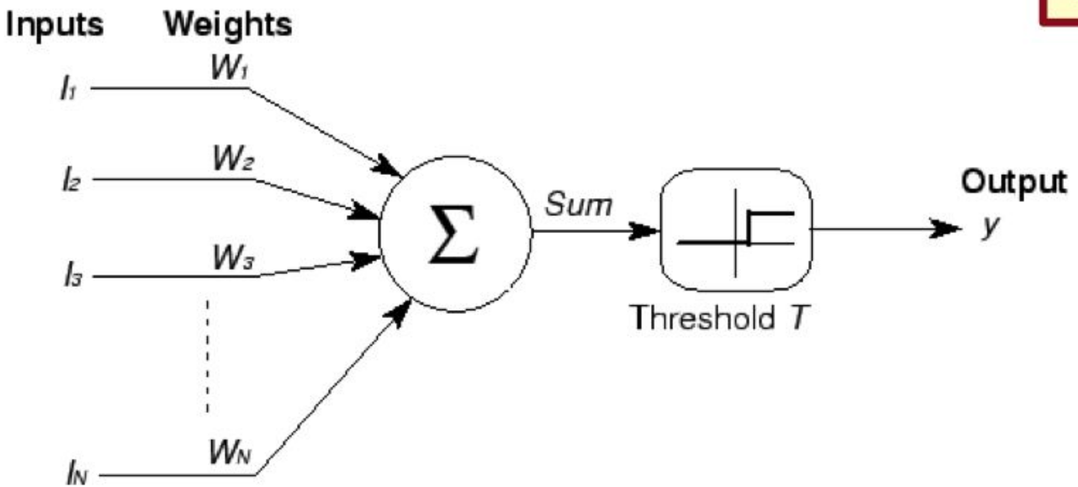
Penn Engineering

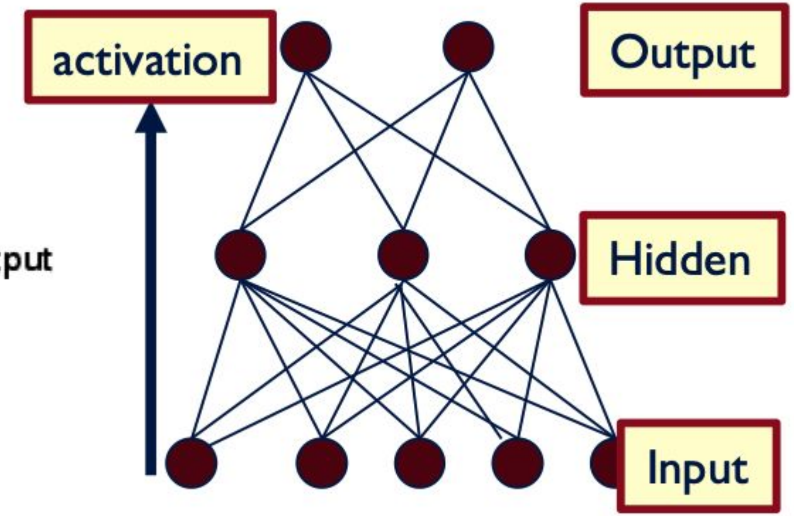
Content

- Neural Networks
- CNN
- Probability Review for Naive Bayes

Part 1: Neural Network

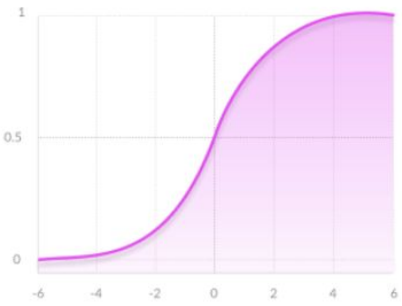
Basics about NN



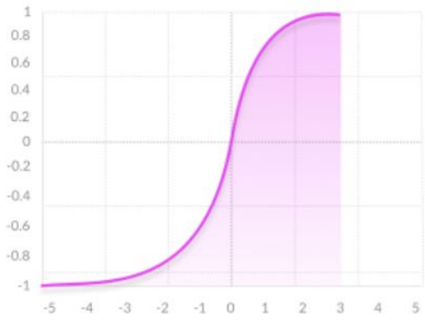


Activation Function

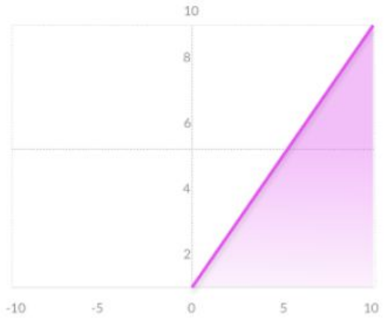
Sigmoid / Logistic



Tanh



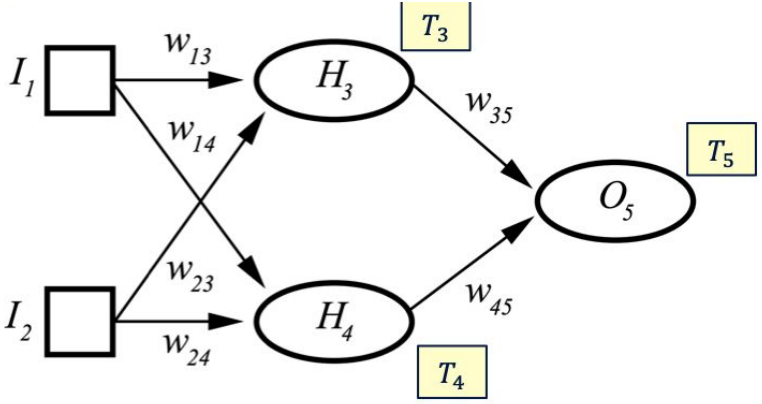
ReLU



If you are working on CNN project, choosing a right activation function might be crucial

More to explore:
https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/

Backpropagation



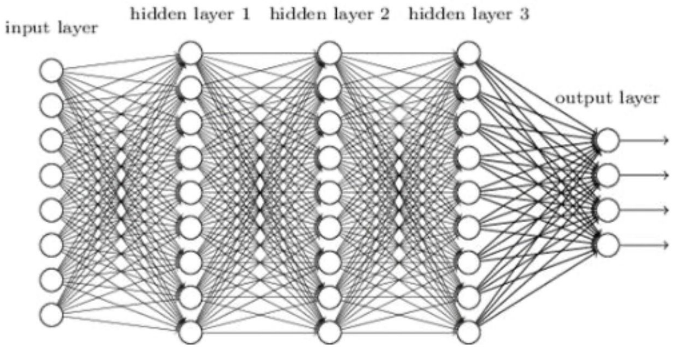

```
for i, data in enumerate(trainloader, 0):
    # get the inputs; data is a list of [inputs, labels]
    inputs, labels = data

    # zero the parameter gradients
    optimizer.zero_grad()

    # forward + backward + optimize
    outputs = net(inputs)
    loss = criterion(outputs, labels)
    loss.backward()
    optimizer.step()
```

Tips for debugging:

- check gradient if you suspect nn is not really actively training

Part 2: CNN

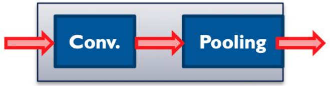
MLP vs CNN



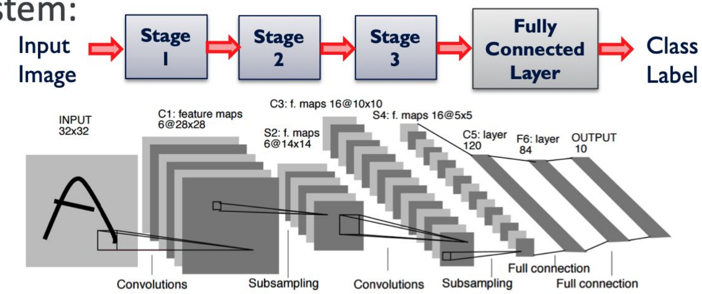
Useful in lot of fields such as:

Reinforcement Learning or topics with relatively small feature space

- One stage structure:



- Whole system:

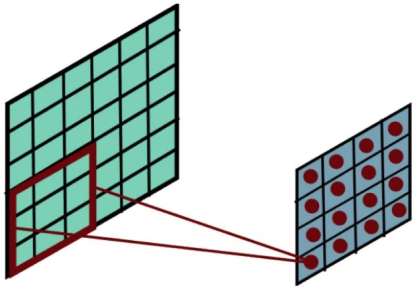


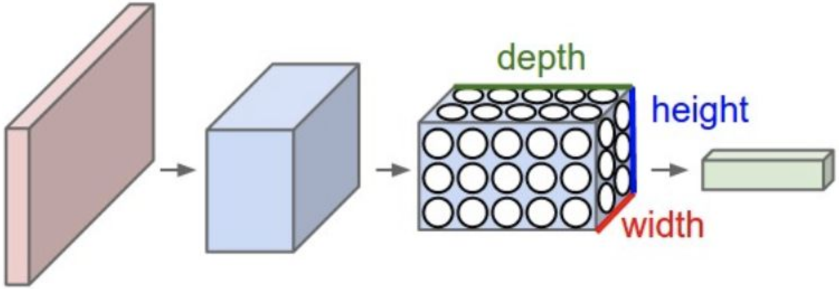
Deal with image/ Image Recognition or Image Generation:

large feature space (eg. 1024*1024*3)

Convolutional Operator

- objective: smartly shrink down feature sizes while maintain information





- Aggregate neighbor information
- Demo: https://cs231n.github.io/convolutional-networks/

Be Careful with Channel

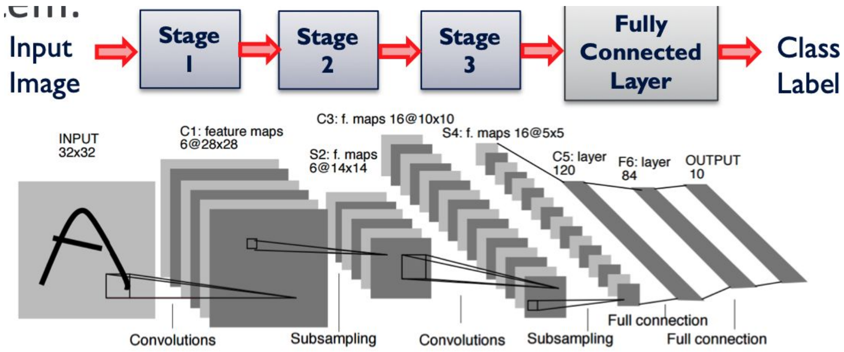
```
class Net(nn.Module):
    def __init__(self):
        super(Net, self).__init__()
        self.conv1 = nn.Conv2d(3, 6, 5)
        self.pool = nn.MaxPool2d(2, 2)
        self.conv2 = nn.Conv2d(6, 16, 5)
        self.fc1 = nn.Linear(16 * 5 * 5, 120)
        self.fc2 = nn.Linear(120, 84)
        self.fc3 = nn.Linear(84, 10)

    def forward(self, x):
        x = self.pool(F.relu(self.conv1(x)))
        x = self.pool(F.relu(self.conv2(x)))
        x = x.view(-1, 16 * 5 * 5)
        x = F.relu(self.fc1(x))
        x = F.relu(self.fc2(x))
        x = self.fc3(x)
        return x
```

CONV2D

CLASS `torch.nn.Conv2d(in_channels: int, out_channels: int, kernel_size: Union[T, Tuple[T, T]], stride: Union[T, Tuple[T, T]] = 1, padding: Union[T, Tuple[T, T]] = 0, dilation: Union[T, Tuple[T, T]] = 1, groups: int = 1, bias: bool = True, padding_mode: str = 'zeros')`

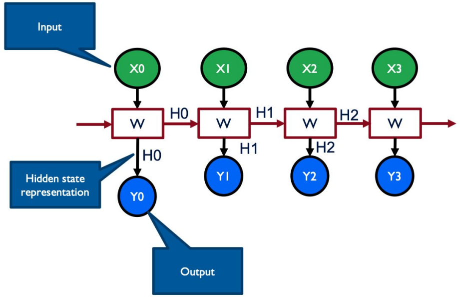
They need to match among layers

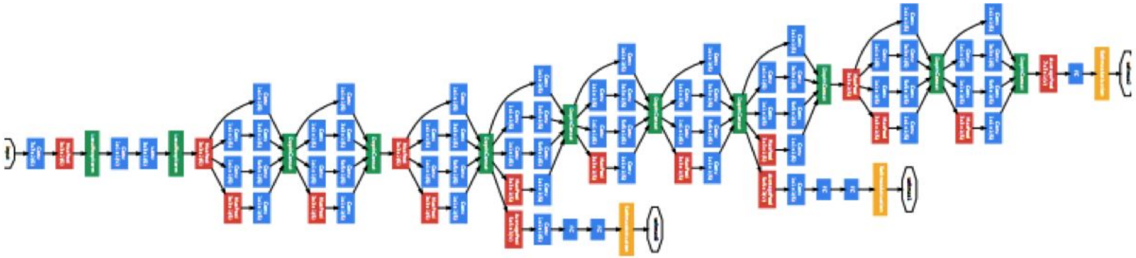
Now Building the Network!



Recurrent Neural Networks

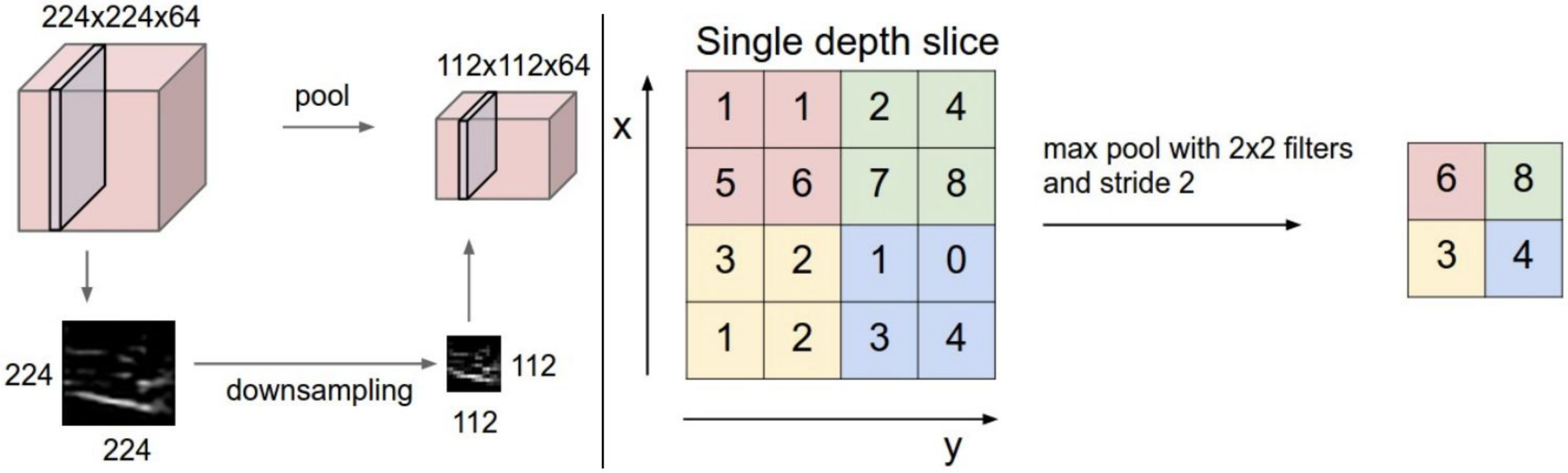
- Infinite uses of finite structure





Pooling

- Another trick to shrink down the image
- eg. Max Pooling - only remain the max one in a certain region



Debug in Colab

import pdb

https://colab.research.google.com/drive/12jFFanVuZgm_wfrCFGBjVigPU2chqZD7?usp=sharing

More Courses

- ESE 546 Principles of Deep Learning (fall)
- CIS 680 Vision & Learning (fall)
- ESE 650 Learning In Robotics (spring) - modules about reinforcement learning
- CIS 522 - Deep Learning for Data Science(spring)

Part 2: Probability for Naive Bayes

Random Variables

Suppose Dan tosses a fair coin 6 times. Examples of random variables (r.v.):

- Let X be the number of tails.
- Let Y be the number of heads.
- Let W be the number of tails in the first 3 throws.
- Let T be the number of tosses until first head.
- Let V denote whether or not the fifth toss is a head.
- Let U be the number of consecutive tail-head tosses.
- ❌ Let M be the probability that the first toss is a head.

R.V.'s are numerical descriptions of an **outcome** of a statistical experiment.

Penn Engineering

Probability & Events

An event is a **set of outcomes** of a statistical experiment. The probability of an event is the chance that the outcome of a trial belongs (or 'lands' on) to the set of outcomes that is the event, i.e. (# of outcomes in event / Total # of outcomes)

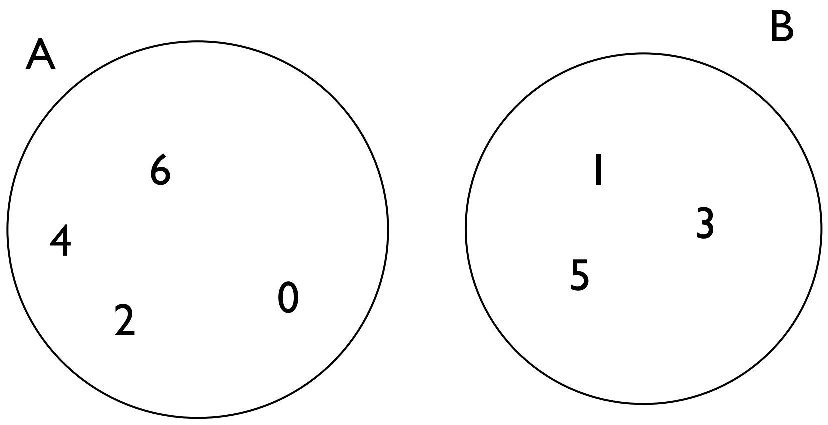
Number of heads (x)	0	1	2	3	4	5	6
P(X=x)	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)	P(X=5)	P(X=6)

Example: Let *X* denote the number of heads that Dan tossed. Find P(*X* is even).

The event "X is even" is the set of values of X = {0, 2, 4, 6}.
P(X is even) = P(X=0) + P(X=2) + P(X=4) + P(X=6)

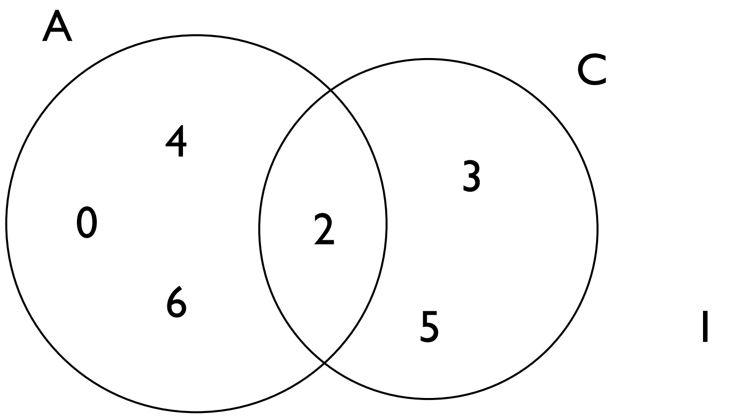
Combining Events: Exclusivity

Mutually-exclusive events



where A is the event where Dan tossed an even number of heads and B odd.

Not exclusive events



where A is the event where Dan tossed an even number of heads and C prime.

Combining Events: Independence

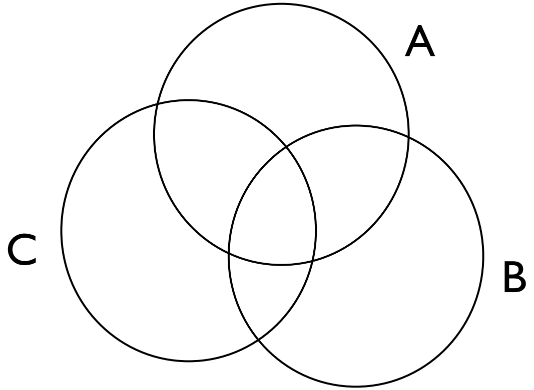
Two events A and B are independent iff the following formula applies:

$$P(A \cap B) = P(A) \cdot P(B)$$

The probability of A occurring **should not** be affected by the occurrence of event B and vice versa.

Principle of Inclusion - Exclusion (PIE)

Let's take a look at an example between events A, B and C:



$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

Expectation

Going back to our example of Dan tossing a fair coin 6 times,...

Number of heads (x)	0	1	2	3	4	5	6
P(X=x)	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)	P(X=5)	P(X=6)
x•P(X=x)	0	P(X=1)	2•P(X=2)	3•P(X=3)	4•P(X=4)	5•P(X=5)	6•P(X=6)

$$E(X) = \sum_{x=0}^{6} x \cdot P(X = x)$$

Expectation of a r.v. X is the **weighted average** of the possible values that X can take, each value being weighted according to the probability of that event occurring.

Variance

Going back to our example of Dan tossing a fair coin 6 times,...

Number of heads (x)	0	1	2	3	4	5	6
P(X=x)	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)	P(X=5)	P(X=6)
x•P(X=x)	0	P(X=1)	2•P(X=2)	3•P(X=3)	4•P(X=4)	5•P(X=5)	6•P(X=6)
x^2•P(X=x)	0	P(X=1)	4•P(X=2)	9•P(X=3)	16•P(X=4)	25•P(X=5)	36•P(X=6)

$$\text{Var}(X) = \sum_{x=0}^{6} x^2 \cdot P(X = x) - (E(X))^2$$

Variance of a r.v. X, informally, measures how far a set of numbers is **spread out from their average value**.

Variance

Going back to our example of Dan tossing a fair coin 6 times,...

Number of heads (x)	0	1	2	3	4	5	6
P(X=x)	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)	P(X=5)	P(X=6)
x•P(X=x)	0	P(X=1)	2•P(X=2)	3•P(X=3)	4•P(X=4)	5•P(X=5)	6•P(X=6)
x^2•P(X=x)	0	P(X=1)	4•P(X=2)	9•P(X=3)	16•P(X=4)	25•P(X=5)	36•P(X=6)

$$\text{Var}(X) = \sum_{x=0}^{6} x^2 \cdot P(X = x) - (E(X))^2$$

Variance of a r.v. X, informally, measures how far a set of numbers is **spread out from their average value**.

Conditional Probability

We denote the probability of an event A occurring, given another event B occurring with the following notation:

$$P(A \mid B)$$

We assume / know that event B is has occurred / is occurring. And with this assumption, we aim to find the probability A occurring. So, if 2 events A and B are independent,

$$P(A \mid B) = P(A); P(B \mid A) = P(B)$$

because the given occurrence does not affect the probability of the event of interest occurring.