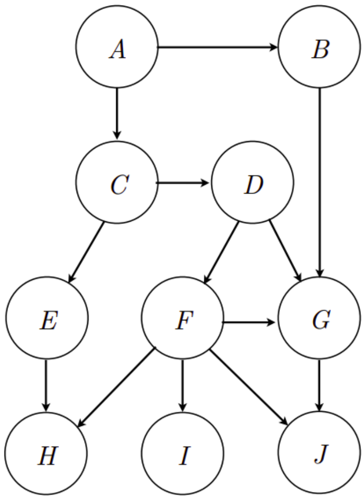
CS-337 and CS-335 Assignment 5

D.Chandra Sekhar
190050031

November 19, 2021

1.1



(a) $C \perp B$?

(b) $C \perp B \mid A$?

(c) $C \perp B \mid A, J$?

(d) $C \perp B \mid A, J, D$?

(e) $C \perp G$?

(f) $C \perp G \mid B$?

(g) $C \perp G \mid B, D$?

(h) $C \perp G \mid B, D, H$?

(i) $C \perp G \mid B, D, H, E$?

(j) $B \perp I \mid J$?

Figure 1: D-Separation Questions

Figure 1: Question 1.1

(I'll be reffering to head-to-head nodes as H-H and Head-to-Tail nodes as H-T and tail-to-tail nodes as T-T). An observation that helps in this question is that nodes H, J always act as H-H nodes irrespective of how they are traversed.

1. Is $C \perp B$?

   Ans: False
   Reason: Consider the path $\mathcal{P} = C \to A \to B$. Here the node A is not blocking as it is a T-T node and it's value is unobserved. Hence this is an unblocked path.

2. Is $C \perp B|A$?

   Ans: True
   Reason: We'll prove it by contradiction. Let $\mathcal{P}$ be an unblocked path from $C \to B$. Notice that this path can't have H, J in it, as they act as unobserved H-H nodes, which block the path. So removing the edges which have H, J in them, we can easily observe that the only way to reach B is by using either $A \to B$ edge or by using $G \leftarrow B$ edge. But in case of $A \to B$ edge, A is an observed non H-H node and hence will block the path. In case $G \leftarrow B$ edge, we must've used either $D \to G$ or $F \to G$ edges to reach G and then use $G \leftarrow B$ edge to reach B. But in either case, G acts as an unobserved H-H node and hence will block the path. So there is no way to reach B via an unblocked path. Hence we have obtained the contradiction.

3. Is $C \perp B|A, J$?

   Ans: False
   Reason: Consider the path $\mathcal{P} = C \to D \to F \to J \to G \to B$. Nodes D, F, G on this path are H-T nodes and their values are unobserved. Node J is a H-H node and it's value is observed. Hence this path is not blocked.

4. Is $C \perp B|A, J, D$?

   Ans: True
   Reason: We'll prove it by contradiction. Let $\mathcal{P}$ be an unblocked path from $C \to B$. Notice that this path can't have H in it, as it acts as unobserved H-H node, which blocks the path. Hence we can't have any edge incident of H in path $\mathcal{P}$. Also we can't use the $A \to B$ edge because A is an observed non H-H node and hence will block the path. and for similar reason we should also not use the $D \to F$ and $D \to G$ edges in our path. Now on removing these edges from the graph, it is trivial to observe that C is disconnected from G in this graph. Hence we have obtained the contradiction.

5. Is $C \perp G$?

   Ans: False
   Reason: Consider the path $\mathcal{P} = C \to D \to G$. The only node D here is H-T node and is unobserved. Hence this path is unblocked.

6. Is $C \perp G|B$?

   Ans: False
   Reason: Consider the path $\mathcal{P} = C \to D \to G$. The only node D here is H-T node and is unobserved. Hence this path is unblocked.

7. Is $C \perp G|B, D$?

   Ans: True
   Reason: We'll prove this by contradiction. Let $\mathcal{P}$ be an unblocked path from $C \to G$. Notice that this path can't have H, J in it, as they act as unobserved H-H nodes, which block the path. Also we can't use $D \to G$ or $B \to G$ edges to reach to G because, they'll cause the path to block as D, B are observed, non H-H nodes. So we need to only use the $F \to G$ edge. But again to reach F from C we need to use the $D \to F$ edge. But again if we inlcude this edge($D \to F$), D will block this path. So there is no way to get to F(and hence to G) via an unblocked path. So we have arrived at the contradiction.

8. Is $C \perp G|B, D, H$?

   Ans: False
   Reason: Consider the path $\mathcal{P} = C \to E \to H \to F \to G$. E, F are H-T and T-T nodes respectively and both values are unobserved. H is a H-H node and it's value is observed. Hence this path is unblocked.

9. Is $C \perp G|B, D, H, E$?

   Ans: True
   Reason: Let us do it by contradiction. Let $\exists$ a path $\mathcal{P}$ from $C \to G$ which is not-blocked. $\mathcal{P}$ can't have node J as it an H-H node and is unobserved. So we should reach G by using either $B \to G, D \to G, F \to G$. But D, B are already observed and non H-H hence will block the path. So we need to get to G using $F \to G$ edge only. But again reaching to F from C is not possible without using the $E \to H$ or $D \to F$ edges. But if we use either of them, then either $EorD$ will act as non H-H, observed nodes and will block the path. So there is no way to reach F(and hence G) via a non-blocking path. Hence we have arrived at the contradiction.

10. Is $B \perp I | J$?

    Ans: False

    Reason: Consider the path $\mathcal{P} = B \to A \to C \to D \to F \to I$. All the nodes on this path are non- H-H nodes and their values are unobserved. Hence this path is unblocked.

1.2

$\texttt{Prob} = \pi_{n \in \texttt{Nodes}} \texttt{P}(n|\texttt{Parents}(n))$.
$= \texttt{P(A)} * \texttt{P(B|A)} * \texttt{P(C|A)} * \texttt{P(D|C)} * \texttt{P(E|C)} * \texttt{P(F|D)} * \texttt{P(G|F,D,B)} * \texttt{P(H|E,F)} * \texttt{P(I|F)} * \texttt{P(J|F,G)}$

1.3

Probability	Parameter-Count
P(A)	$2^0 = 1$
P(B|A)	$2^1 = 2$
P(C|A)	$2^1 = 2$
P(D|C)	$2^1 = 2$
P(E|C)	$2^1 = 2$
P(F|D)	$2^1 = 2$
P(G|F,D,B)	$2^3 = 8$
P(J|F,G)	$2^2 = 4$
P(I|F)	0(since I = F)
P(H|E,F)	0(since H = E OR F)
Joint-Prob	23

Number of parameters required for $\texttt{P}(n|\texttt{Parents(n)})$ is $= 2^{|\texttt{Parents(n)}|}$.

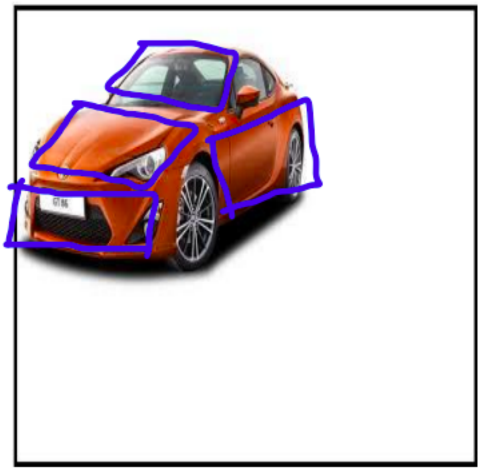
This is because for each setting of each Parent to a particular value in $\{0, 1\}$, we need a bernoulli-parameter. For random variables I, H we don't need any parameters because we already know their dependence on their parents.

The number of parameters of Joint-Prob is just the summation of number of parameters required for each of its components.

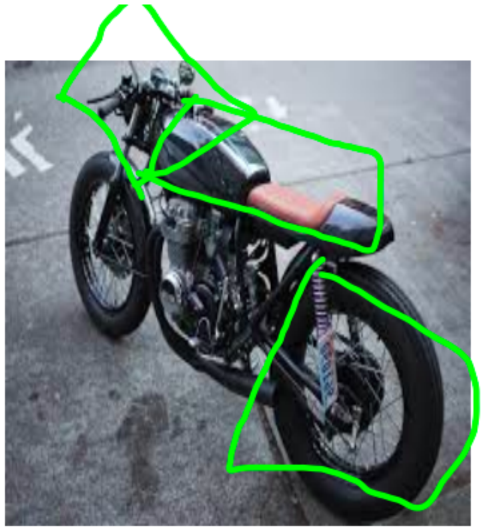
Question-2

Task-1

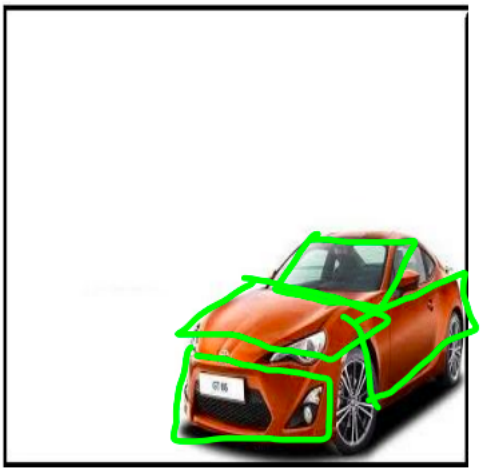
The regions which are used to distinguish between the car and bike are marked by bounding rectangles in the following figures:



(a) Figure3.1



(b) Figure3.2



(c) Figure3.3

Basically it's the door, numberplate, glass pane, engine region of car that makes it distinct. And for bikes it is the handle, seat, big tyres that make it unique.

- **Object Location**: Because of the weight sharing, the convolution operation is equivariant. i.e. conv(shift(f)) = shift(conv(f)). Hence this makes the CNN's robust to object **translations** .

- **Scaling**: CNNs in general are not extremely robust to scale changes. We deliberately need to train CNNs with scaled images(via data-augmentation) to make them robust. Assuming the scale changes are not very drastic, the CNNs can be robust to an extent to scale changes. Because of the convolutional layers and Maxpool layers acting **locally**, CNNs first capture the local properties like edges, corners etc. and then they combine these information in the deeper convolutional layers to capture even bigger/global properties like glasspane region, numberplate region ..etc. So even on increase in scale, if the neural net is deep enough, our CNN will be able to combine these local properties into a larger/global properties(like glass-pane ..etc.) eventually. Hence will be able to correctly identify the object. **So in short the *localness* of conv, max-pool layers and the depth of the network provides robustness to scale changes**.

.

Task2

Instead of using a softmax at the last layer, we can use the sigmoid function for **each of the output**. We declare all the classes which have a sigmoid value/score > a `predefined-threshold` as belonging to that image. We need to do this because, the softmax inherently imposes the condition that sum-of-probabilities of the image belonging to various classes as = 1. But in case of image having multiple objects, this is no longer the case. The sigmoid doesnot impose this condition. Hence it allows for image having more than one-object and hence will help us in this task.

The Limitations of this solution are the following:

- You can't have as many convoultion layers as you had earlier. We need to decrease them. This is because in the deep convolutional layers(i.e. the ones closer to output layer) the features for different objects can start falling within the same patch and hence may lead to bad classification. Hence we need to reduce the number of convolution layers.
- The interdependence between various class labels is not being taken into account incase of final layer being a sigmoid layer. Though this is not a problem if we have a large dataset, it is a problem if we have a limited dataset. Also getting a bigger dataset in case of Multi-Label-Classification is more difficult because the number of annotations per image is now > 1.

Task-3

- The filter sizes(in both convolution and max-pool layers) needs to be reduced so as to ensure that, pixels from different objects are not falling within the same patch. And also Lesser number of colvolutional layers are to be used because of the same reason.
- Instead now we need to have more fully-connected layers as they help in modelling more complex object interactions.

The Limitations of this approach is(are):

- The increased computational and data requirements because of the fully-connected layers.
- Due to occulsion some times only some features of the object might be visible. Ex: Only a single tyre of a bike being visible, only the handle of the bike being visible. Our model doesnot explicitly try to infer things from partially observed features.