

Project 5 Questions

Instructions

- 3 questions.
- Write code where appropriate; feel free to include images or equations.
- **We do NOT expect you to fill up each page with your answer.** Some answers will only be a few sentences long, and that is okay.

Questions

Q1:

- (a) Explain these common terms in machine learning in your words:
 - (i) Bias
 - (ii) Variance
- (b) Define these terms in the context of evaluating a classifier:
 - (i) Overfitting
 - (ii) Underfitting
- (c) Does bias and variance have any impact on overfitting and underfitting. Can you describe a brief real-world scenario where you can find/observe this?

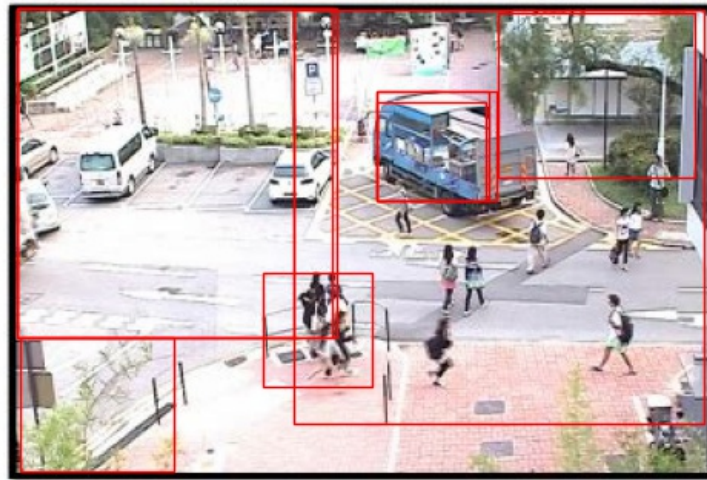
Please answer overleaf.

A1: Your answer here.

- (a)
 - (i) Bias measures the difference between predicted model and true model.
 - (ii) Variance is the difference between the output values of the models trained on different training datasets.
- (b)
 - (i) Overfitting: The model fits the training set well, but performs poorly on the test set.
 - (ii) Underfitting: the model cannot fit the training set.
- (c) Yes, overfitting has large variance while underfitting has large bias.

For example, when a person wants to sell a house, he investigates the house price according to the size of the house and fits a straight line. But it is often not able to sell at the highest price, because there are many other factors, such as geographic location, etc., which is underfitting with large bias.

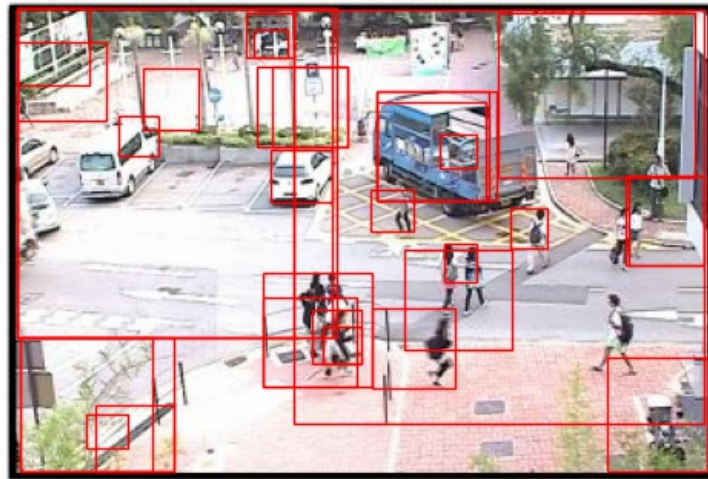
Q2: Suppose you had to test the selective search algorithm on an image from pedestrian detection dataset (for example: an image taken from traffic camera as shown below), do you think that selective search algorithm will suggest bounding boxes over person(s).



- (a) If yes, what is the justification in your words for this successful behavior of the algorithm? If not, then can you think in which cases does it fail? Can you suggest one way to improve or modify the approach?

A2: Your answer here.

1. I think no. The goal of selective search algorithm is detect objects at any scale. The region of pedestrian is relative small. We can decrease limited size of region to capture small scale like following figure.



Q3: If you were to apply selective search algorithm to detect interesting regions for skin diseases, how do you think the algorithm would have performed?

- (a) Which among the four similarity (color, texture, size and shape) do you think contributes more to automatic detection of interesting regions in this case?
- (b) Combining what you understood so far, what do you think “objectness” means and do you agree that selective search algorithm inherently finds regions of interest based on “objectness”?

Please answer overleaf.

A3: Your answer here.

- (a) In this case, I think the weight of the color will contribute more, because skin disease usually reddens with inflammation. Second maybe the texture, especially some skin diseases can cause damage to the skin surface.
- (b) Objectness tells you how likely a certain image window is to contain an "object" (as opposed to uncharacterizable background).

Yes, I agree.

Feedback? (Optional)

Please help us make the course better. If you have any feedback for this assignment, we'd love to hear it!