# 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)

(ii)

(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 is an error from erroneous assumptions in the learning algorithm

(ii) variance is an error from sensitivity to small fluctuations in the training set.

(b) (i) Overfitting:Overfitting refers to a modeling error that occurs when a function corresponds too closely to a dataset

(ii) Underfitting: Underfitting refers to a model that can neither model the training dataset nor generalize to new dataset.

(c) Yes, High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).
Variance is the algorithm's tendency to learn random things irrespective of the real signal by fitting highly flexible models that follow the error/noise in the data too closely (overfitting).
Exam overfitting - When you study for an exam, only by practicing questions from previous years' exams. You then discover to your horror that xx% of this year's questions are new, and you get a much lower score than on your practice ones.

**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.

(a) No, the core of objective detection is how to effectively remove redundant candidate regions. In fact, most of the redundant candidate regions overlap. Selective search uses this to merge adjacent overlapping regions from the bottom up to reduce redundancy. Selective search has fast speed and high recall rate.
However, the regions of each pedestrian are too small, which might probably be filtered away.
We can decrease the threshold of filtering small regions to adjust the pedestrian's region size, but the number of proposal regions might explosively grow.

**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) color and texture might contribute more. For color, the surface color of disease region usually will be quite different from surroundings,like melanin. If not, the skin tissues can be implemented skin pathological staining. For texture, the surface textures of disease region are quite unique, like scaly texture, which would be a useful feature representation.

(b) Objectness means choice based on objective facts, such as color, texture, etc.
Yes, selective search algorithm inherently finds regions of interest based on ”objectness”

## Feedback? (Optional)

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