Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

EEU44C16-1/EE5C16-1

FACULTY OF SCIENCE, TECHNOLOGY, ENGINEERING AND MATHEMATICS

SCHOOL OF ENGINEERING

Electronic and Electrical Engineering

Engineering **Semester 1, 2022**

Senior Sophister/M.A.I./M.Sc.

Deep Learning and its Applications

Final Exam

15th Dec 2022 **RDS Simmonscourt** **09:30-11:30**

Dr François Pitié

Instructions to candidates:

Answer all questions.
For the MCQ questions (Q1-24), there is no negative marks or partial marks given for incorrect or partially correct answers. The given marks will be 2.5 for fully correct answers and 0 otherwise.
Make sure that the MCQ questions are answered in the provided MCQ answer sheet. Only non-MCQ questions are allowed in the scripts.
Return both the filled-in answer sheet and your scripts.

Materials Permitted for this Examination:

Calculators

Additional information:

Supporting Material for Gradient Computations

Assuming $\mathbf{a}, \mathbf{b}, \mathbf{A}$ are independent of $\mathbf{w}$, below is a list of useful gradient computations:

$$\frac{\partial \mathbf{a}^\top \mathbf{w}}{\partial \mathbf{w}} = \mathbf{a}$$

$$\frac{\partial \mathbf{b}^\top \mathbf{A}\mathbf{w}}{\partial \mathbf{w}} = \mathbf{A}^\top \mathbf{b}$$

$$\frac{\partial \mathbf{w}^\top \mathbf{A}\mathbf{w}}{\partial \mathbf{w}} = (\mathbf{A} + \mathbf{A}^\top)\mathbf{w} \qquad \text{(or } 2\mathbf{A}\mathbf{w} \text{ if } A \text{ symmetric)}$$

$$\frac{\partial \mathbf{w}^\top \mathbf{w}}{\partial \mathbf{w}} = 2\mathbf{w}$$

$$\frac{\partial\, \mathbf{a}^\top \mathbf{w}\mathbf{w}^\top \mathbf{b}}{\partial\, \mathbf{w}} = (\mathbf{a}\mathbf{b}^\top + \mathbf{b}\mathbf{a}^\top)\mathbf{w}$$

Exam Number: _____

Seat Number : _____

Mark your exam number below

0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9

All your MCQ answers must be filled in on this answer page.

For **True** or **False** questions, mark Ⓣ or Ⓕ. For questions with multiple choices, mark all solutions that are correct (for instance Ⓐ Ⓑ Ⓒ Ⓓ).

***Q*.1.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.2.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
***Q*.3.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
***Q*.4.** Ⓐ Ⓑ Ⓒ Ⓓ
***Q*.5.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.6.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
***Q*.7.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.8.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.9.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.10.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.11.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.12.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.13.** Ⓐ Ⓑ Ⓒ Ⓓ
***Q*.14.** Ⓐ Ⓑ Ⓒ
***Q*.15.** Ⓐ Ⓑ Ⓒ Ⓓ
***Q*.16.** Ⓐ Ⓑ Ⓒ Ⓓ
***Q*.17.** Ⓐ Ⓑ Ⓒ
***Q*.18.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.19.** Ⓐ Ⓑ Ⓒ Ⓓ
***Q*.20.** Ⓐ Ⓑ Ⓒ Ⓓ
***Q*.21.** Ⓐ Ⓑ Ⓒ
***Q*.22.** Ⓐ Ⓑ Ⓒ Ⓓ
***Q*.23.** Ⓐ Ⓑ Ⓒ Ⓓ
***Q*.24.** Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ
***Q*.25.** essay question
***Q*.26.** essay question
***Q*.27.** essay question
***Q*.28.** essay question
***Q*.29.** essay question
***Q*.30.** essay question
***Q*.31.** essay question
***Q*.32.** essay question
***Q*.33.** essay question
***Q*.34.** essay question

This page was intentionally left blank.

Q.1. Which of the following models with input x_1, x_2, parameters w_1, w_2 and noise $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, are linear in the parameters and can be used for Least Squares? (mark all models that are linear)

Ⓐ $y = \log(w_1 + x_1) + w_2 x_2 + \varepsilon$

Ⓑ $y = w_2 + w_1 x_1 + {w_1}^2 x_1 + {w_1}^3 x_1 + {w_1}^4 x_1 + \varepsilon$

Ⓒ $y = w_1 x_1 + w_2 x_2 + x_2^3 + \varepsilon$

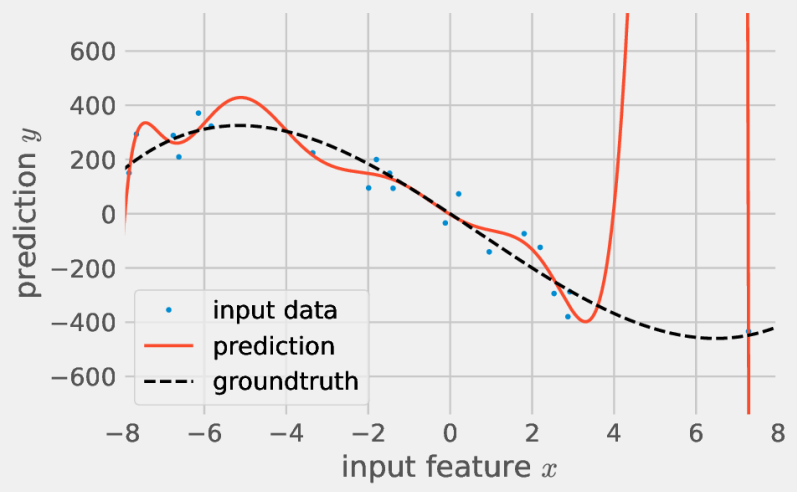
Ⓓ $y = \exp(x_1 w_1 + w_2 + \varepsilon)$

Ⓔ $y = \sin(x_1) w_1 + \cos(x_2) w_2 + 3\varepsilon$

Ⓕ $y = x_1 w_1 + w_2 \varepsilon$

[2.5 marks]

Q.2. We are applying a regression to fit a polynomial to the following dataset. The predicted model does not fit the data very well. Why? (mark all answers that explain reasonably the situation)



Ⓐ there is overfitting

Ⓑ there is underfitting

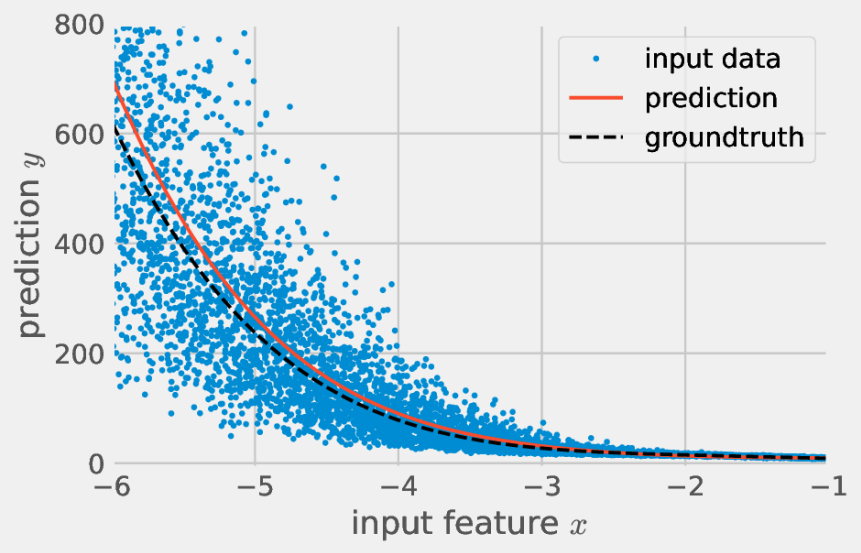
Ⓒ the training set is too small

Ⓓ the testing set is too small

Ⓔ the error is not independent and Gaussian

[2.5 marks]

Q.3. We are applying a regression to fit a polynomial to the following dataset. The predicted model does not fit the data very well. Why? (mark all answers that explain reasonably the situation)



Ⓐ there is overfitting

Ⓑ there is underfitting

Ⓒ the training set is too small

Ⓓ the testing set is too small

Ⓔ the error is not independent and Gaussian

[2.5 marks]

Q.4. Which of the following facts about logistic regression are true? (multiple choices possible):

Ⓐ The logistic model is estimated by way of Maximum Likelihood estimation.

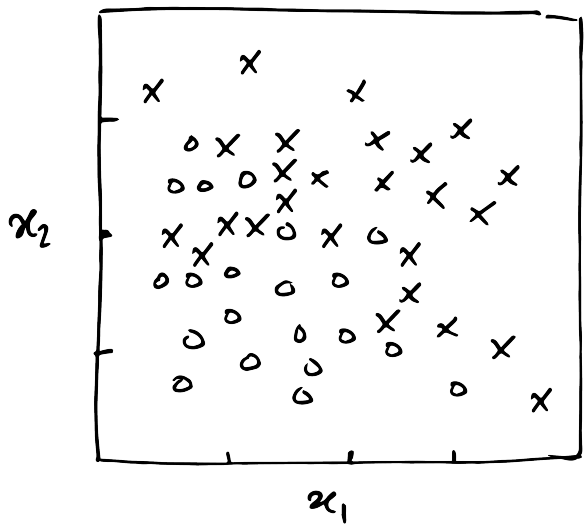
Ⓑ The Loss for both Linear SVM and Logitstic Regression are the same.

Ⓒ The range of logit values is 0 to 1

Ⓓ A small value of the negative log-likelihood statistic indicate that the statistical model is a good fit of the data.

[2.5 marks]

Q.5. Consider the binary class dataset below, with 2 features (x_1, x_2) and 2 classes (cross and circle). Which of the following classification techniques is (are) apriori well suited to this dataset? (choose all that are suitable)



- Ⓐ Logistic Regression
- Ⓑ Random Forest
- Ⓒ SVM with RBF kernel with small γ
- Ⓓ SVM with RBF kernel with large γ
- Ⓔ 10-Nearest-Neighbour
- Ⓕ 3-Nearest-Neighbour

[2.5 marks]

Q.6. Which of the following binary classifiers can produce smooth decision boundary maps? (select all methods that produce decision boundaries that are not jagged/rough)

- Ⓐ 5-nearest neighbours
- Ⓑ Adaboost or Random Forest
- Ⓒ SVM with RBF kernel
- Ⓓ SVM with polynomial kernel
- Ⓔ Logistic Regression

[2.5 marks]

Q.7. Consider a binary classifier with the following confusion matrix:

Confusion Matrix	actual:0	actual:1
predicted: 0	TN=49	FN=10
predicted: 1	FP=0	TP=41

What is the Recall? (one of the proposed solutions is correct)

Ⓐ 0.00%
Ⓑ 80.39%
Ⓒ 83.05%
Ⓓ 89.13%
Ⓔ 90.00%
Ⓕ 100.00%

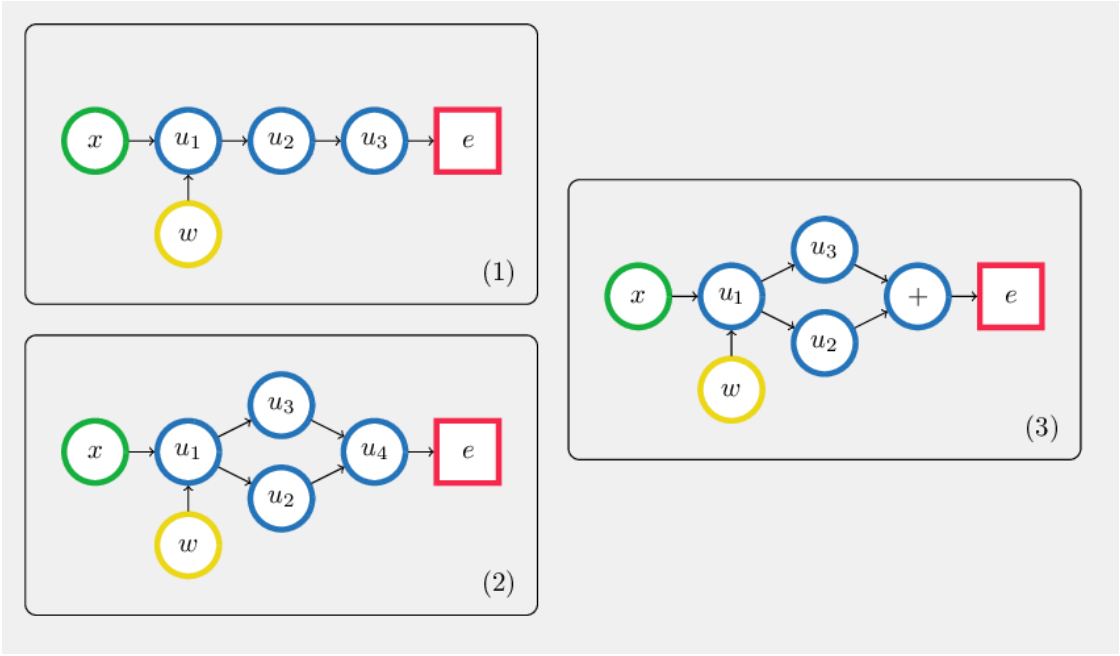
[2.5 marks]

Q.8. In which of the following architectures, does weight sharing occur? (multiple choices possible)

Ⓐ convolutional neural Network
Ⓑ recurrent neural network.
Ⓒ fully connected neural network
Ⓓ ResNet
Ⓔ GoogLeNet
Ⓕ VGG

[2.5 marks]

Q.9. Rank the following networks according to how likely they are to suffer from vanishing gradient. For instance, (1) < (2) < (3) means that (3) is the most likely to suffer from vanishing gradient, and (1) the least likely. u_1, u_2, u_3, u_4 are artifical neurons with sigmoid activations. The unit marked with + is simply an addition.



Ⓐ (1) < (2) < (3)
Ⓑ (1) < (3) < (2)
Ⓒ (2) < (1) < (3)
Ⓓ (2) < (3) < (1)
Ⓔ (3) < (1) < (2)
Ⓕ (3) < (2) < (1)

[2.5 marks]

Q.10. Based on the following code, what is the shape of the output?

```
inputs = keras.layers.Input(shape=(25, 11, 5))
out = keras.layers.Conv2D(3,
        kernel_size=(5,5),
        strides=(2, 2),
        padding="same")(inputs)
model = keras.models.Model(inputs, out)
```

Ⓐ $11 \times 4 \times 3$

Ⓑ $12 \times 5 \times 3$

Ⓒ $13 \times 6 \times 3$

Ⓓ $14 \times 7 \times 3$

Ⓔ $23 \times 9 \times 3$

Ⓕ $25 \times 11 \times 3$

[2.5 marks]

Q.11. Based on the same code as in previous question, what is the *number of parameters* of the model?

Ⓐ 25

Ⓑ 28

Ⓒ 125

Ⓓ 128

Ⓔ 375

Ⓕ 378

[2.5 marks]

Q.12. Suppose we have a 14-layers deep CNN which takes 1 day to train on a GPU. At test time, it takes 50ms to evaluate a single data point. We change the architecture by adding a dropout layer after the first and second layers (with rates of 0.2 and 0.3 respectively). What would be new compute time to evaluate a single test data point with this new architecture?

Ⓐ less than 25ms.
Ⓑ less than 50ms.
Ⓒ exactly 50ms.
Ⓓ more than 50ms.
Ⓔ more than 100ms.
Ⓕ not enough information to say.

[2.5 marks]

Q.13. Which of the following statements about Batch Normalisation (BN) are correct?

Ⓐ BN is used as a layer.
Ⓑ with BN, we need to increase the dropout.
Ⓒ BN is useful for domain adaptation.
Ⓓ BN rescales its input values so as to have 0 mean and variance 1.

[2.5 marks]

Q.14. Which of the following statements are correct? (multiple choices possible)

Ⓐ In Auto-Encoders, the distribution of the latent variable is the normal distribution.
Ⓑ In Variational Auto-Encoders, the distribution of the latent variable is regularised to the normal distribution.
Ⓒ In GANs, the input distribution of the noise variable can be the normal distribution.

[2.5 marks]

Q.15. We want to do some transfer learning. Having access to a large amount of data for the target application, why is fine-tuning necessary?

Ⓐ It results in optimized weights based on the distribution of the target data.
Ⓑ It results in optimized weights based on the distribution of the original data.
Ⓒ It removes unnecessary weights.
Ⓓ It merges the original and target data resulting in a better model.

[2.5 marks]

Q.16. Which of the following statements are correct? (mark all the statements that are true)

Ⓐ LSTM, by Hochreiter et Schmidhuber, was proposed in 2008.
Ⓑ ResNet, by He et al., uses recurrent layers.
Ⓒ GoogLeNet, by Szegedy et al., uses inception modules.
Ⓓ The Cifar10 dataset consists of about 60,000 32x32 colour images in 10 classes (airplane, automobile, etc.).

[2.5 marks]

Q.17. We want to classify 144 characters long tweets using a simple RNN network. Which of the following statements are correct. (select all true answers)

```
1 inputs = Input(shape=(144, 56))
2 h = SimpleRNN(10, return_sequences=False)(inputs)
3 k = SimpleRNN(10, return_sequences=True)(h)
4 output = Dense(3, activation='softmax')(k)
5 m = Model(inputs, output)
```

Ⓐ The part of the code (lines 2 and 3) which defines the RNN is incorrect.
Ⓑ The classification involves 10 classes
Ⓒ The loss function is not defined but, apriori, it could be L_1 or L_2.

[2.5 marks]

Q.18. For the listing below, which of these statements are true? (multiple choices possible)

```
(x_train, y_train), (x_test, y_test) = load_data()
x_train = sequence.pad_sequences(x_train, maxlen=400)
x_test = sequence.pad_sequences(x_test, maxlen=400)
model = Sequential()
model.add(Embedding(input_dim=5000, output_dim=50, input_length=400))
model.add(Dropout(0.2))
model.add(Conv1D(filters=250, kernel_size=3, padding='valid', \
          activation='relu', strides=2))
model.add(GlobalMaxPooling1D())
model.add(Dense(250))
model.add(Dropout(0.2))
model.add(Activation('relu'))
model.add(Dense(1))
model.add(Activation('sigmoid'))
model.compile(loss='binary_crossentropy', optimizer='adam')
model.fit(x_train, y_train, batch_size=32, epochs=2)
```

Ⓐ The network is an Autoencoder.

Ⓑ The model is a FeedForward Neural Net.

Ⓒ The model is used for classification.

Ⓓ The model input tensor in an image.

Ⓔ The model archictecture uses Recurrent Layers.

Ⓕ The model archictecture uses Convolutional Layers.

[2.5 marks]

Q.19. We have the following loss expression:

$$E(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{w}^\top \mathbf{x}_i - \mathbf{y}_i\|^2 + \lambda \sum_{j=1}^{P} w_i \log(w_i)$$

What does the term $\sum_{j=1}^{P} w_i \log(w_j)$ refers to? (mutiple choices possible)

Ⓐ a regularisation term

Ⓑ a discriminator loss

Ⓒ a cross-entropy loss

Ⓓ the negative log likelihood

[2.5 marks]

Q.20. Below is a Keras implementation of VGG16.

```
# Block 1
x = Conv2D(64, (3, 3), activation='relu', padding='same', name='block1_conv1')(img_input)
x = Conv2D(64, (3, 3), activation='relu', padding='same', name='block1_conv2')(x)
x = MaxPooling2D((2, 2), strides=(2, 2), name='block1_pool')(x)
# Block 2
x = Conv2D(128, (3, 3), activation='relu', padding='same', name='block2_conv1')(x)
x = Conv2D(128, (3, 3), activation='relu', padding='same', name='block2_conv2')(x)
x = MaxPooling2D((2, 2), strides=(2, 2), name='block2_pool')(x)
# Block 3
x = Conv2D(256, (3, 3), activation='relu', padding='same', name='block3_conv1')(x)
x = Conv2D(256, (3, 3), activation='relu', padding='same', name='block3_conv2')(x)
x = Conv2D(256, (3, 3), activation='relu', padding='same', name='block3_conv3')(x)
x = MaxPooling2D((2, 2), strides=(2, 2), name='block3_pool')(x)
# Block 4
x = Conv2D(512, (3, 3), activation='relu', padding='same', name='block4_conv1')(x)
x = Conv2D(512, (3, 3), activation='relu', padding='same', name='block4_conv2')(x)
x = Conv2D(512, (3, 3), activation='relu', padding='same', name='block4_conv3')(x)
x = MaxPooling2D((2, 2), strides=(2, 2), name='block4_pool')(x)
# Block 5
x = Conv2D(512, (3, 3), activation='relu', padding='same', name='block5_conv1')(x)
x = Conv2D(512, (3, 3), activation='relu', padding='same', name='block5_conv2')(x)
x = Conv2D(512, (3, 3), activation='relu', padding='same', name='block5_conv3')(x)
x = MaxPooling2D((2, 2), strides=(2, 2), name='block5_pool')(x)
# Classification block
x = Flatten(name='flatten')(x)
x = Dense(4096, activation='relu', name='fc1')(x)
x = Dense(4096, activation='relu', name='fc2')(x)
x = Dense(classes, activation='softmax', name='predictions')(x)
```

Ⓐ The output tensor values are all positive or null.

Ⓑ The tensor width and height are larger than in `block2\_conv2` than in `block3\_conv1`.

Ⓒ The predicted tensor size is a vector of size 4096×1

Ⓓ the loss function could be chosen as L_2.

[2.5 marks]

Q.21. If we used one-hot word encoding for an application summarising books, what would be the size of the input vectors of your NLP neural system?

Ⓐ about 128 (the number of characters)

Ⓑ typically about a few 1000s.

Ⓒ about 100,000

[2.5 marks]

Q.22. Which of the following facts are true? (select all statements that are true)

Ⓐ L_2 regularisation helps setting some parameters to zero and thus can be used for pruning the parameters of a model.

Ⓑ L_1 regularisation helps setting some parameters to zero and thus can be used for pruning the parameters of a model.

Ⓒ using L_1 helps with training convergence.

Ⓓ using L_2 helps with training convergence.

[2.5 marks]

Q.23. Which of the following statements are correct? (select all the statements that are true)

Ⓐ One issue with Transformers is that they are not simple to parallelise

Ⓑ One issue with RNNs is that they are not simple to parallelise

Ⓒ The computational complexity of a Self-Attention layer is quadratic on the size of the input dimension

Ⓓ Transformers follows an encoder-decoder structure

[2.5 marks]

Q.24. What is the depth of the following network?

```
nfeatures = 32
model = keras.Sequential(
  [
    keras.Input(shape=(6, nfeatures)),
    layers.SimpleRNN(64),
    layers.Dense(1, activation="sigmoid"),
  ]
)
```

Ⓐ 3

Ⓑ 6

Ⓒ 7

Ⓓ 32

Ⓔ 64

Ⓕ 65

[2.5 marks]

Q.25. We have the following loss expression:

$$E(\mathbf{w}) = \exp\left(-\left(\mathbf{w}^\top \mathbf{x} - 1\right)^2\right)$$

Add a L_2 regularisation term, derive the corresponding expression of the gradient and write down the steps of the gradient descent applied to the minimisation of this loss. **[4 marks]**

Q.26. What gives non-linearity to a neural network? Why is it required? What is/are the potential problems associated with it? **[4 marks]**

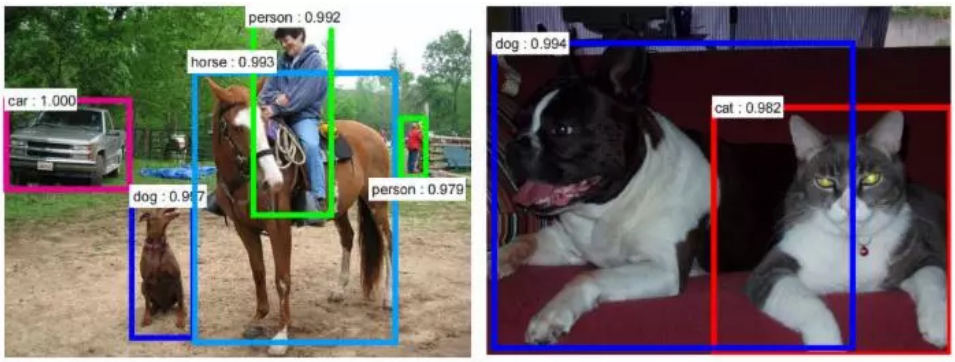
Q.27. A company has developed a new tool for detecting tax fraud. The company claims an 94.67% accuracy and impresses the investors. Comment. **[4 marks]**

Q.28. Give 3 examples CNN architectures that take a fixed-size input tensor. Does a CNN for classification necessarily require a fixed-size input tensor? Explain. **[4 marks]**

Q.29. We have timeseries for indicators of an engine, given by various temperature and pressure sensors. The sensors are read every second. We have collected data over the lifetime of 150 engines.

Design a deep learning system that can predict potential anomalies in the operation of the engine. **[4 marks]**

Q.30. Object detection results are often depicted as in Figure below. Explain the precise signification of the reported numbers and what the underlying model used is.



[4 marks]

Q.31. Explain how an object classifier like VGG can be tuned to reach a particular level of sensitivity? **[4 marks]**

Q.32. Explain what a *Deep Neural Network*, a *unit*, a *neuron* and a *layer* are. **[4 marks]**

Q.33. Explain the problem of vanishing gradients and name 5 techniques/architectures/components that can be used to mitigate this issue, and, for each of these methods, explain briefly in one single line how the method helps with the problem of vanishing gradients. **[4 marks]**

Q.34. Explain what Auto-Encoders are and what is the objective behind Variational Auto-Encoders. **[4 marks]**