**International Journal of Engineering and Advanced Technology (IJEAT)**

**ISSN: 2249 – 8958, Volume-8, Issue-3S, February 2019**

# Marks Prediction Using Linear Regression

## Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam 

***ABSTRACT--- This report discusses the use of linear regression for predicting exam marks. The report includes a literature survey on the use of AI algorithms for predicting exam marks, and presents a system architecture and design for predicting marks using linear regression. The methodology for the project is also discussed, including data collection, feature selection, model selection, and prediction. The report includes a code implementation of linear regression for predicting exam marks in Python, and evaluates the performance of the model using mean squared error, root mean squared error, and R-squared score. The report concludes by discussing potential future enhancements to the system. Overall, this report provides a comprehensive overview of the use of linear regression for predicting exam marks and serves as a useful reference for researchers and practitioners in the field of education.***

**I. INTRODUCTION** Linear regression is a widely used statistical technique that involves predicting a continuous output variable based on one or more input variables. It is a powerful tool for analyzing and modeling data in various fields, including economics, finance, and healthcare. In recent years, with the emergence of machine learning, linear regression has become an essential technique for developing predictive models. The purpose of this research paper is to provide an overview of linear regression in machine learning, including its techniques, merits, and demerits.

In the first paragraph of the introduction, the concept of linear regression is introduced, and its importance in various fields is highlighted. The second paragraph emphasizes the increasing relevance of linear regression in machine learning. The third paragraph sets out the purpose of the research paper, which is to provide an overview of linear regression in machine learning. Finally, the fourth paragraph outlines the structure of the research paper.

### II. LITERATURE SURVEY

Linear regression has been extensively studied in the literature, and numerous techniques and variations have been proposed to improve its performance. One of the most popular techniques is ordinary least squares (OLS) linear regression, which involves minimizing the sum of the squared differences between the predicted values and the

### actual values in the dataset. Other techniques include ridge regression, which adds a penalty term to the OLS objective function to prevent overfitting, and lasso regression, which adds a penalty term that encourages sparse solutions.

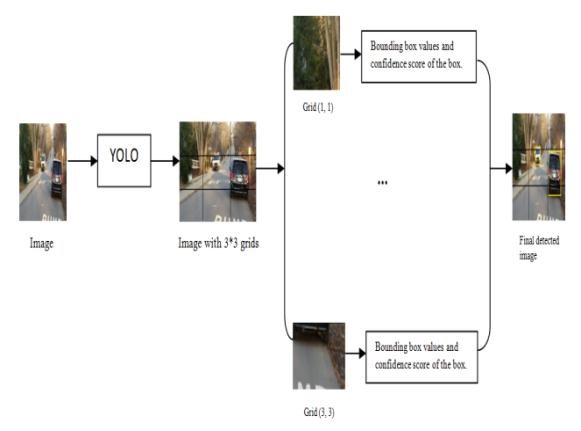
In the first paragraph of the literature review, the importance of linear regression in machine learning is emphasized. The second paragraph provides an overview of ordinary least squares linear regression, which is the most widely used technique in linear regression. The third paragraph highlights the other techniques used to improve the performance of

### 

### linear regression. The fourth paragraph discusses the merits of linear regression, including its simplicity and interpretability, and its ability to handle non-linear relationships. The fifth paragraph discusses the demerits of linear regression, including its assumption of linearity, and its sensitivity to outliers.

### III. WORKING OF YOLO ALGORITHM

First, an image is taken and YOLO algorithm is applied. In our example, the image is divided as grids of 3x3 matrixes. We can divide the image into any number grids, depending on the complexity of the image. Once the image is divided, each grid undergoes classification and localization of the object. The objectness or the confidence score of each grid is found. If there is no proper object found in the grid, then the objectness and bounding box value of the grid will be zero or if there found an object in the grid then the objectness will be 1 and the bounding box value will be its corresponding bounding values of the found object. The bounding box prediction is explained as follows. Also, Anchor boxes are used to increase the accuracy of object detection which also explained below in detail.



### Figure 1: Working of YOLO

*3.1. Bounding box predictions:*

YOLO algorithm is used for predicting the accurate bounding boxes from the image. The image divides into S x S grids by predicting the bounding boxes for each grid and class probabilities. Both image classification and object localization techniques are applied for each grid of the image and each grid is assigned with a label. Then the algorithm checks each grid separately and marks the label which has an object in it and also marks its bounding boxes.

The labels of the gird without object are marked as zero.

*Published By:*

*Blue Eyes Intelligence Engineering*

*Retrieval Number: C11240283S19/19©BEIESP & Sciences Publication*

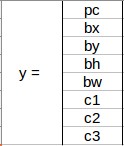
### REAL-TIME OBJECT DETECTION WITH YOLO



### Figure 2: Example image with 3x3 grids

Consider the above example, an image is taken and it is divided in the form of 3 x 3 matrixes. Each grid is labelled and each grid undergoes both image classification and objects localization techniques. The label is considered as Y.

Y consists of 8 values.



### Figure 3: Elements of label Y

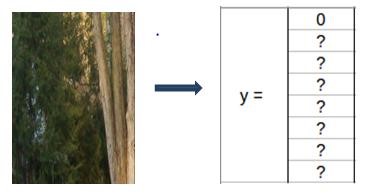
Pc – Represents whether an object is present in the grid or not. If present pc=1 else 0. bx, by, bh, bw – are the bounding boxes of the objects (if

present). c1, c2, c3 – are the classes. If the object is a car then c1

and c3 will be 0 and c2 will be 1.

In our example image, the first grid contains no proper

object. So it is represented as,

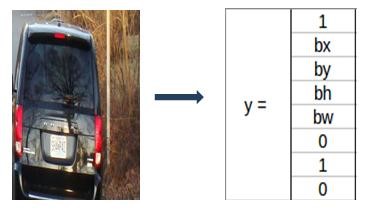


**Figure 4: Bounding box and Class values of grid 1.**

In this grid, there exists no proper object so the pc value is 0.

And rest of the values are doesn’t matter because there exist no object. So, it is represented as ?.

Consider a grid with the presence of an object. Both 5th and 6th grid of the image contains an object. Let’ consider the 6th grid, it is represented as.



**Figure 5: Bounding box and Class values of grid 6.**

In this table, 1 represents the presence of an object. And bx, by, bh, bw are the bounding boxes of the object in the 6th grid. And the object in that grid is a car so the classes are (0,1,0). The matrix form of Y in this is Y=3x3x8.

For the 5th grid also the matrix will be little similar with different bounding boxes by depending on the objects position in the corresponding grid.

If two or more grids contain the same object then the center point of the object is found and the grid which has that point is taken. For this, to get the accurate detection of the object we can use to methods. They are Intersection over Union and Non-Max Suppression. In IoU, it will takes the actual and predicted bounding box value and calculates the

IoU of two boxes by using the formulae, IoU = Area of Intersection / Area of Union.

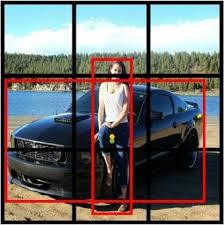
If the value of IoU is more than or equal to our threshold value (0.5) then it's a good prediction. The threshold value is just an assuming value. We can also take greater threshold value to increase the accuracy or for better prediction of the object.

The other method is Non-max suppression, in this, the high probability boxes are taken and the boxes with high IoU are suppressed. Repeat this until a box is selected and consider that as the bounding box for that object.

*3.2. ACCURACY IMPROVEMENT*

**ANCHOR BOX:**

By using Bounding boxes for object detection, only one object can be identified by a grid. So, for detecting more than one object we go for Anchor box.



### Figure 6: An example image for anchor box

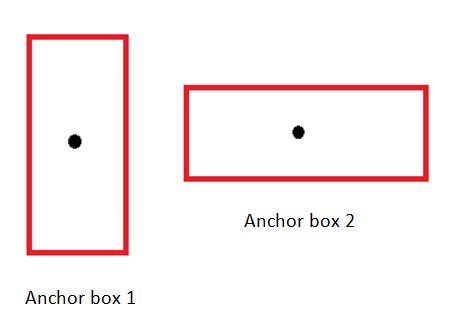
*Published By:*

*Blue Eyes Intelligence Engineering*

*Retrieval Number: C11240283S19/19©BEIESP & Sciences Publication*

### International Journal of Engineering and Advanced Technology (IJEAT)

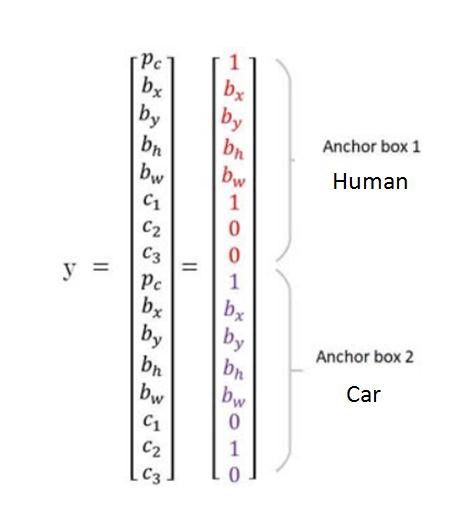
Consider the above picture, in that both the human and the car’s midpoint come under the same grid cell. For this case, we use the anchor box method. The red color grid cells are the two anchor boxes for those objects. Any number of anchor boxes can be used for a single image to detect multiple objects. In our case, we have taken two anchor boxes.



### Figure 7: Anchor boxes

The above figure represents the anchor box of the image we considered. The vertical anchor box is for the human and the horizontal one is the anchor box of the car.

In this type of overlapping object detection, the label Y contains 16 values i.e, the values of both anchor boxes.



### Figure 8: Anchor box prediction values

Pc in both the anchor box represents the presence of the object.

bx, by, bh, bw in both the anchor box represents their

corresponding bounding box values.

The value of the class in anchor box 1 is (1, 0, 0) because the detected object is a human.

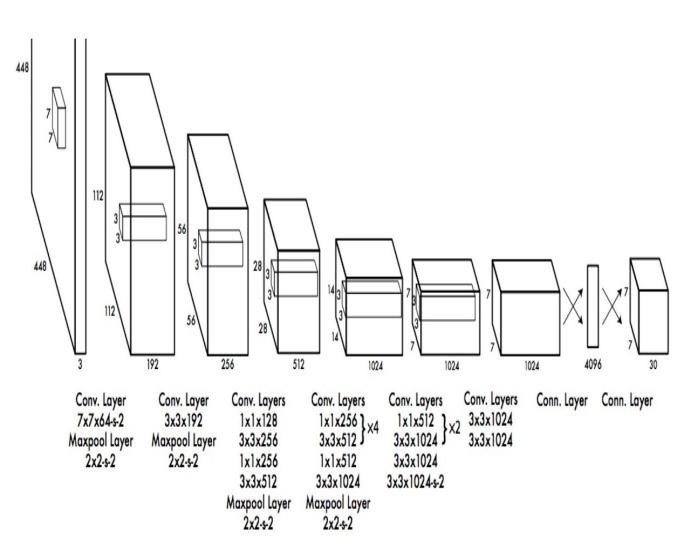
In the case of anchor box 2, the detected object is a car so the class value is (0, 1, 0).

In this case, the matrix form of Y will be Y= 3x3x16 or Y= 3x3x2x8. Because of two anchor box, it is 2x8.

### IV. RESULTS & DISCUSSIOS

The idea of YOLO is to make a Convolutional neural network to predict a (7, 7, 30) tensor.It uses a Convolutional neural network to scale back the spatial dimension to 7x7 with 1024 output channels at every location. By using two **ISSN: 2249 – 8958, Volume-8, Issue-3S, February 2019**

fully connected layers it performs a linear regression to create a 7x7x2 bounding box prediction. Finally, a prediction is made by considering the high confidence score of a box.



### Figure 9: CNN Network Design

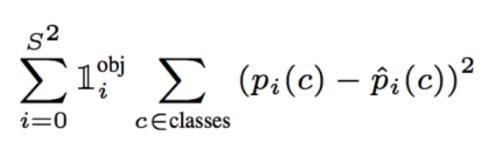
*4.1. Loss function of YOLO algorithm:*

For a single grid cell, the algorithm predicts multiple bounding boxes. To calculate the loss function we use only one bounding box for object responsibility. For selecting one among the bounding boxes we use the high IoU value. The box with high IoU will be responsible for the object.

Various loss functions are:

* Classification loss function
* Localization loss function
* Confidence loss function

Localization loss means the error between the ground truth value and predicted boundary box. Confidence loss is the objectness of the box. Classification loss calculated as, the squared error of the class conditional probabilities for each class:



### Equation 1:Conditional probabilities for each class

Where, in Equation 1, If it is 1 means the object appears in

the cell, or else it is 0.

 is the conditional class probability for class c.

The localization loss is the measure of errors in the predicted boundary box locations and the sizes. The box which is responsible for the object is only counted.

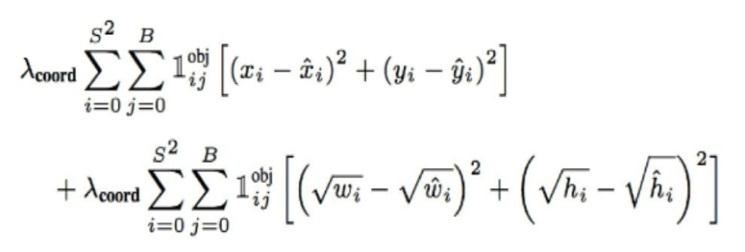


*Published By:*

*Blue Eyes Intelligence Engineering*

*Retrieval Number: C11240283S19/19©BEIESP & Sciences Publication*

### REAL-TIME OBJECT DETECTION WITH YOLO



### Equation 2: The localization loss

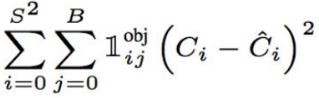
Where,

in Equation 2, Is 1, if the jth bounding box of cell i

is responsible for detecting the object. Otherwise, it is 0.

Increase the weight for the loss of bounding box coordinates.

The Confidence loss, if the object is found in a box the confidence loss is,

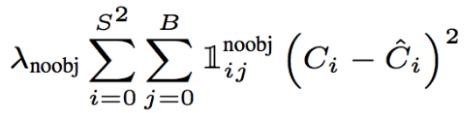


### Equation 3:The Confidence loss

Where, in Equation 3, Is the confidence score of the box j in cell i.

Is 1 if the jth bounding box of cell i is responsible for detecting the object. Otherwise, it is 0.

If the object is not detected then the confidence loss will be,



### Equation 4: Confidence loss if object not detected

Where, in Equation 4, Is the complement of.

Is the confidence score of box j in cell i.



Is the weights down the loss when detecting the background.

### V. CONCLUSION

In this paper, we proposed about YOLO algorithm for the purpose of detecting objects using a single neural network.This algorithm is generalized, it outperforms different strategies once generalizing from natural pictures to different domains. The algorithm is simple to build and can be trained directly on a complete image.Region proposal strategies limit the classifier to a particular region. YOLO accesses to the entire image in predicting boundaries. And also it predicts fewer false positives in background areas.Comparing to other classifier algorithms this algorithm is much more efficient and fastest algorithm to use in real time.

### REFERENCES

1. Joseph Redmon, Santosh Divvala, Ross Girshick, “You Only Look Once: Unified, Real-Time Object Detection”,The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
2. YOLO Juan Du1,”Understanding of Object Detection

Based on CNN Family”,New Research, and Development Center of Hisense, Qingdao 266071, China.

1. Matthew B. Blaschko Christoph H. Lampert, “Learning to Localize Objects with Structured Output Regression”, Published in Computer Vision – ECCV 2008 pp 2-15. 4. Wei Liu, Dragomir Anguelov, Dumitru Erhan, “SSD: Single Shot MultiBox Detector”, Published in Computer Vision – ECCV 2016 pp 21-37.
2. Lichao Huang, Yi Yang, Yafeng Deng, Yinan Yu

DenseBox, “Unifying Landmark Localization with End to End Object Detection”, Published in Computer Vision and Pattern Recognition (cs.CV).

1. Dumitru Erhan, Christian Szegedy, Alexander Toshev,

“Scalable Object Detection using Deep Neural Networks”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2147-2154.

1. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Published in Advances in Neural Information Processing Systems 28 (NIPS 2015).
2. Joseph Redmon, Ali Farhadi,“YOLO9000: Better, Faster, Stronger”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7263-7271.
3. Jifeng Dai, Yi Li, Kaiming He, Jian Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks”, published in: Advances in Neural Information Processing Systems 29 (NIPS 2016).
4. Karen Simonyan, Andrew Zisserman, “Very Deep

Convolutional Networks for Large-Scale Image Recognition”, published in Computer Vision and Pattern Recognition (cs.CV).

*Published By:*

*Blue Eyes Intelligence Engineering*

*Retrieval Number: C11240283S19/19©BEIESP & Sciences Publication*