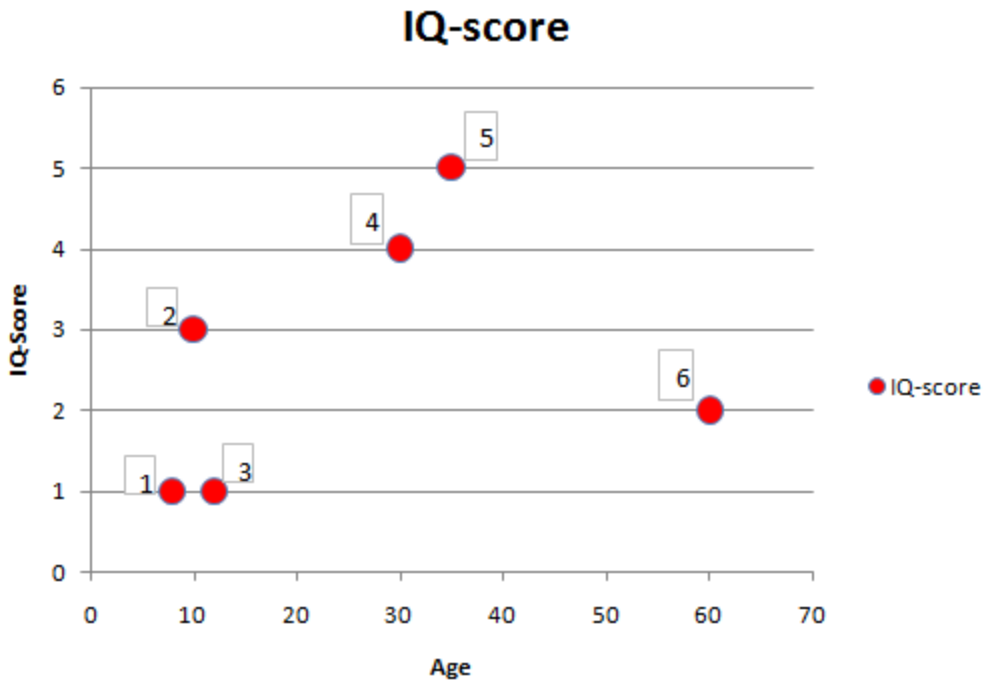
# Gradient Boosting for Regression And classification

Md. Golam Rabiul Alam

Associate Professor

# Gradient Boosting for Regression Example

| Observations | Age | Place of living | IQ-score |
|---|---|---|---|
| 1 | 8 | Village | 1 |
| 2 | 10 | City | 3 |
| 3 | 12 | Village | 1 |
| 4 | 30 | Village | 4 |
| 5 | 35 | City | 5 |
| 6 | 60 | City | 2 |



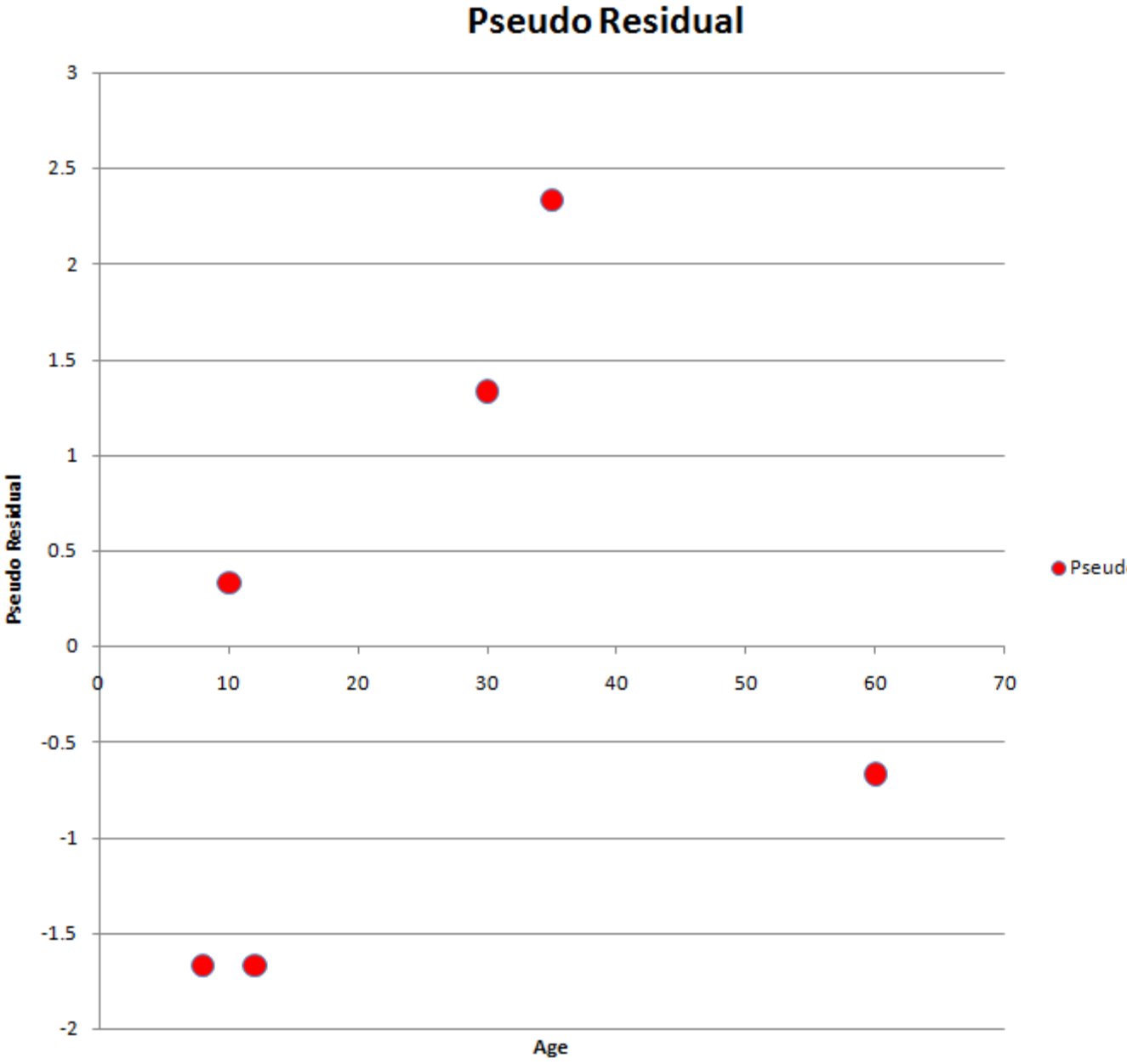
# Gradient Boosting for Regression Example

| Observations | | Age | Place of living | IQ-score | Average IQ Score | Pseudo Residual |
|---|---|---|---|---|---|---|
| | 1 | 8 | Village | 1 | 2.6666667 | -1.66667 |
| | 2 | 10 | City | 3 | | 0.33333 |
| | 3 | 12 | Village | 1 | | -1.66667 |
| | 4 | 30 | Village | 4 | | 1.33333 |
| | 5 | 35 | City | 5 | | 2.33333 |
| | 6 | 60 | City | 2 | | -0.66667 |

| Observations | | Age | Place of living | Pseudo Residual |
|---|---|---|---|---|
| | 1 | 8 | Village | -1.66667 |
| | 2 | 10 | City | 0.33333 |
| | 3 | 12 | Village | -1.66667 |
| | 4 | 30 | Village | 1.33333 |
| | 5 | 35 | City | 2.33333 |
| | 6 | 60 | City | -0.66667 |

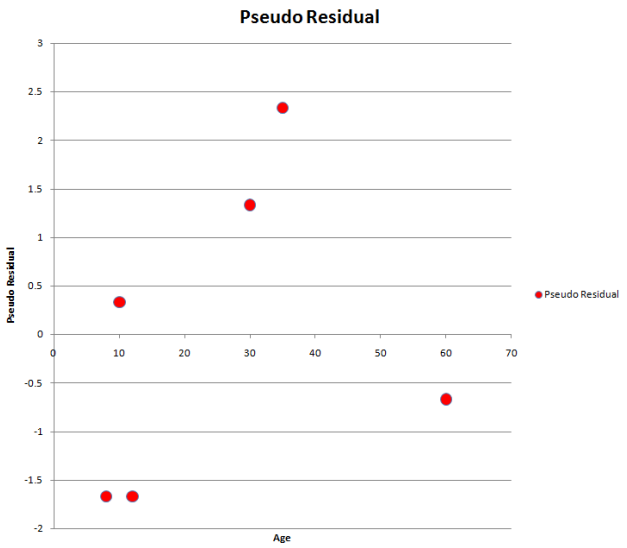
# Gradient Boosting for Regression Example

| | Age | Average of consecutive two persons age |
|---|---|---|
| 1 | 8 | |
| 2 | 10 | 9 |
| 3 | 12 | 11 |
| 4 | 30 | 21 |
| 5 | 35 | 32.5 |
| 6 | 60 | 47.5 |



# Gradient Boosting for Regression Example

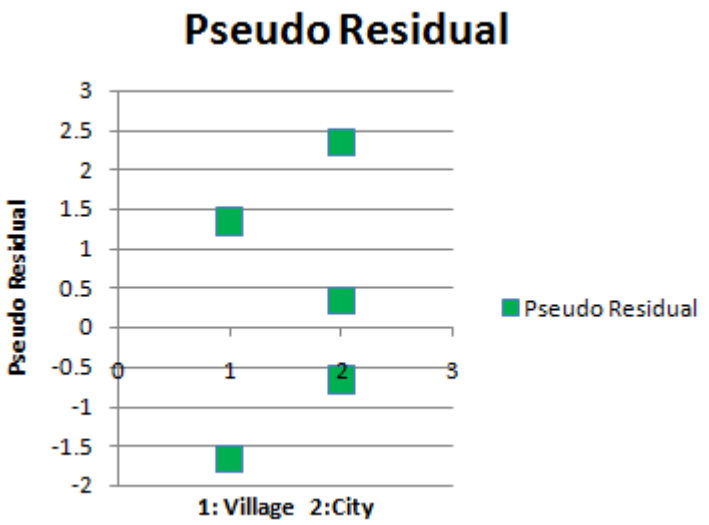
| | Age | Pseudo Residual (PR) | | Average PR score of Left side instances | Average PR score of Right side instances | Sum of square residuals (SSR)/2 |
|---|---|---|---|---|---|---|
| 1 | 8 | -1.66667 | Considering age 9 as the separator | -1.66667 | 0.33333 | 5 |
| 2 | 10 | 0.33333 | Considering age 11 as the separator | -0.66667 | 0.33333 | 6 |
| 3 | 12 | -1.66667 | Considering age 21 as the separator | -1.000003333 | 0.9999967 | 3.66666667 |
| 4 | 30 | 1.33333 | Considering age 32.5 as the separator | -0.41667 | 0.83333 | 5.625 |
| 5 | 35 | 2.33333 | Considering age 47.5 as the separator | 0.13333 | -0.66667 | 6.4 |
| 6 | 60 | -0.66667 | | | | |
| | | | | Minimum SSR/2= | 3.6666667 | |



# Gradient Boosting for Regression Example

| Observations | Age | Place of living | Pseudo Residual |
|---|---|---|---|
| 1 | 8 | Village | -1.66667 |
| 2 | 10 | City | 0.33333 |
| 3 | 12 | Village | -1.66667 |
| 4 | 30 | Village | 1.33333 |
| 5 | 35 | City | 2.33333 |
| 6 | 60 | City | -0.66667 |

| Place of living | Pseudo Residual |
|---|---|
| Village | -1.66667 |
| City | 0.33333 |
| Village | -1.66667 |
| Village | 1.33333 |
| City | 2.33333 |
| City | -0.66667 |



# Gradient Boosting for Regression Example

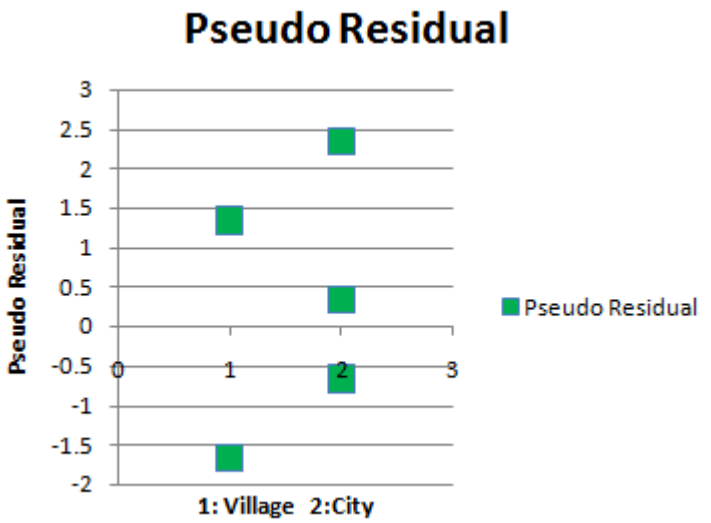
| Place of living | Encoded Place of living | Pseudo Residual (PR) |
|---|---|---|
| Village | 1 | -1.66667 |
| City | 2 | 0.33333 |
| Village | 1 | -1.66667 |
| Village | 1 | 1.33333 |
| City | 2 | 2.33333 |
| City | 2 | -0.66667 |

Considering 'Place of living=Village' as the separator

| Average PR score of Left side instances | Average PR score of Right side instances | Sum of square residuals (SSR) | SSR/2 |
|---|---|---|---|
| -0.66667 | 0.6666633 | 10.6666667 | 5.3333333 |

Minimum SSR/2= 5.3333333

| Place of living | Pseudo Residual |
|---|---|
| Village | -1.66667 |
| City | 0.33333 |
| Village | -1.66667 |
| Village | 1.33333 |
| City | 2.33333 |
| City | -0.66667 |

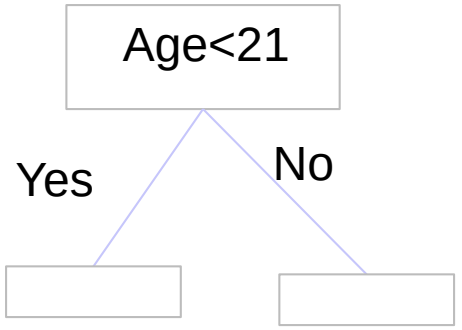


# Gradient Boosting for Regression Example

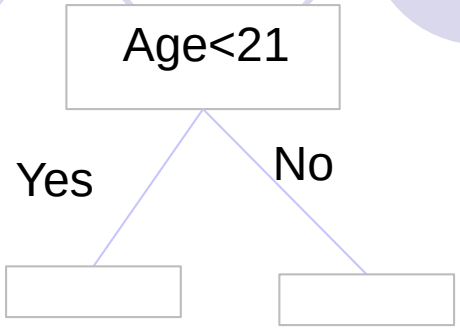
| | Age | Pseudo Residual (PR) | | Average PR score of Left side instances | Average PR score of Right side instances | Sum of square residuals (SSR)/2 |
|---|---|---|---|---|---|---|
| 1 | 8 | -1.66667 | Considering age 9 as the separator | -1.66667 | 0.33333 | 5 |
| 2 | 10 | 0.33333 | Considering age 11 as the separator | -0.66667 | 0.33333 | 6 |
| 3 | 12 | -1.66667 | Considering age 21 as the separator | -1.000003333 | 0.9999967 | 3.66666667 |
| 4 | 30 | 1.33333 | Considering age 32.5 as the separator | -0.41667 | 0.83333 | 5.625 |
| 5 | 35 | 2.33333 | Considering age 47.5 as the separator | 0.13333 | -0.66667 | 6.4 |
| 6 | 60 | -0.66667 | | | | |
| | | | | Minimum SSR/2= | 3.6666667 | |

**Minimum SSR/2 for Age=3.67**

**Minimum SSR/2 Place of Living=5.33**



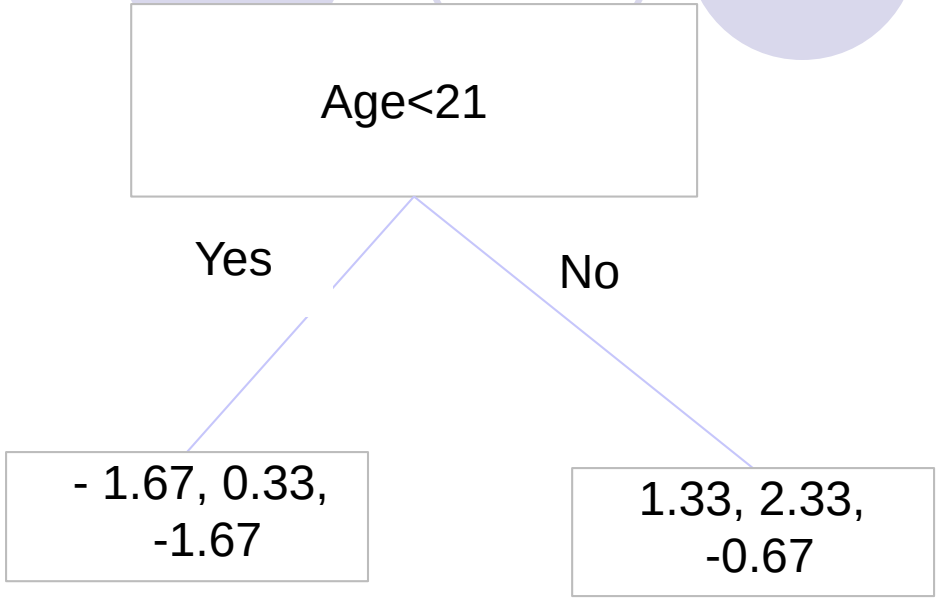
# Gradient Boosting for Regression Example

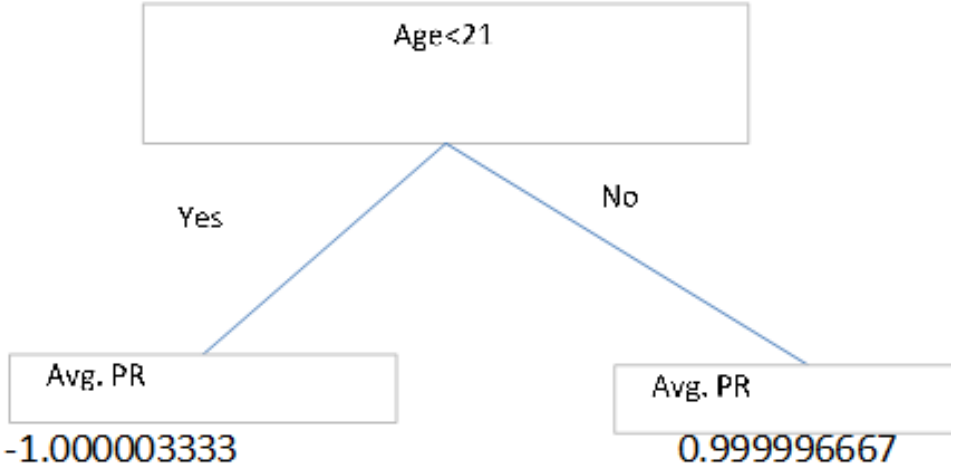



| Observations | | Age | Place of living | Pseudo Residual (PR) |
|---|---|---|---|---|
| | 1 | 8 | Village | -1.66667 |
| | 2 | 10 | City | 0.33333 |
| | 3 | 12 | Village | -1.66667 |

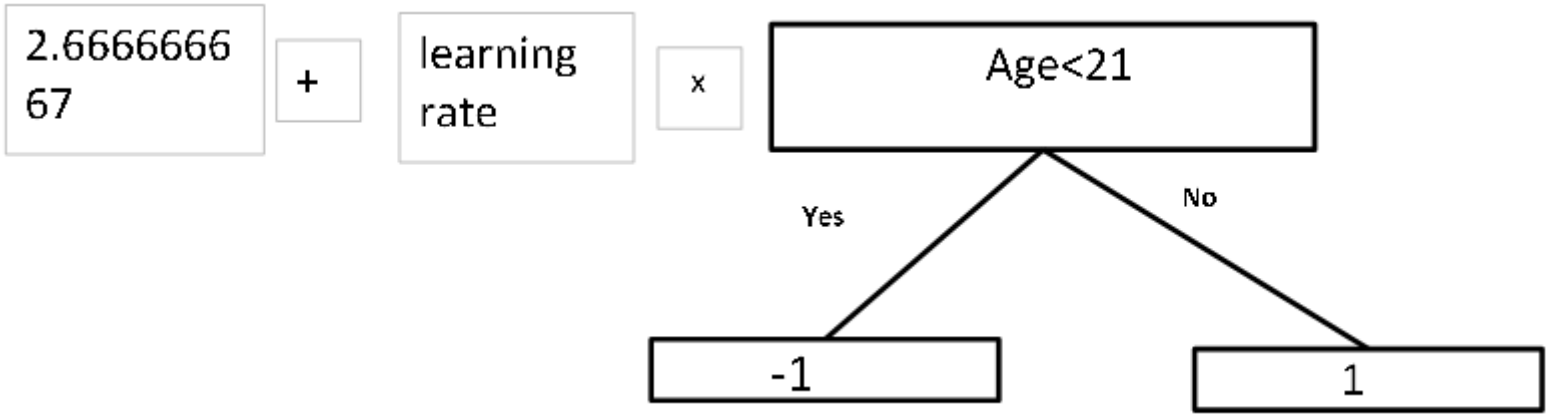
# Gradient Boosting for Regression Example

| | Age | Pseudo Residual (PR) | |
|---|---|---|---|
| 1 | 8 | -1.66667 | Considering age 9 as the separator |
| 2 | 10 | 0.33333 | Considering age 11 as the separator |
| 3 | 12 | -1.66667 | Considering age 21 as the separator |
| 4 | 30 | 1.33333 | Considering age 32.5 as the separator |
| 5 | 35 | 2.33333 | Considering age 47.5 as the separator |
| 6 | 60 | -0.66667 | |



# Gradient Boosting for Regression Example




| Observations | | Age | Place of living | IQ-score | Average IQ Score | Pseudo Residual 1 | Predicted IQ Score | Pseudo Residual 2 |
|---|---|---|---|---|---|---|---|---|
| | 1 | 8 | Village | 1 | | -1.66667 | 2.5666667 | -1.566667 |
| | 2 | 10 | City | 3 | | 0.33333 | 2.5666667 | 0.4333333 |
| | 3 | 12 | Village | 1 | 2.666666667 | -1.66667 | 2.5666667 | -1.566667 |
| | 4 | 30 | Village | 4 | | 1.33333 | 2.7666667 | 1.2333333 |
| | 5 | 35 | City | 5 | | 2.33333 | 2.7666667 | 2.2333333 |
| | 6 | 60 | City | 2 | | -0.66667 | 2.7666667 | -0.766667 |

# Gradient Boosting for Regression Example

| Observations | Age | Place of living | Pseudo Residual 2 |
|---|---|---|---|
| 1 | 8 | Village | -1.566666667 |
| 2 | 10 | City | 0.433333333 |
| 3 | 12 | Village | -1.566666667 |
| 4 | 30 | Village | 1.233333333 |
| 5 | 35 | City | 2.233333333 |
| 6 | 60 | City | -0.766666667 |

| | Age | Pseudo Residual 2 (PR2) | | Average PR2 score of Left side instances | Average PR2 score of Right side instances | Sum of square residuals (SSR)/2 |
|---|---|---|---|---|---|---|
| 1 | 8 | -1.566666667 | Considering age 9 as the separator | -1.56666667 | 0.31333333 | 4.624 |
| 2 | 10 | 0.433333333 | Considering age 11 as the separator | -0.56666667 | 0.28333333 | 5.615 |
| 3 | 12 | -1.566666667 | Considering age 21 as the separator | -0.9 | 0.9 | 3.666666667 |
| 4 | 30 | 1.233333333 | Considering age 32.5 as the separator | -0.36666667 | 0.73333333 | 5.29 |
| 5 | 35 | 2.233333333 | Considering age 47.5 as the separator | 0.153333333 | -0.76666667 | 5.744 |
| 6 | 60 | -0.766666667 | | | | |
| | | | | | Minimum SSR/2= | 3.66666667 |

# Gradient Boosting for Regression Example

| Place of living | Encoded Place of living | Pseudo Residual 2 (PR2) |
|---|---|---|
| Village | 1 | -1.566666667 |
| City | 2 | 0.433333333 |
| Village | 1 | -1.566666667 |
| Village | 1 | 1.233333333 |
| City | 2 | 2.233333333 |
| City | 2 | -0.766666667 |

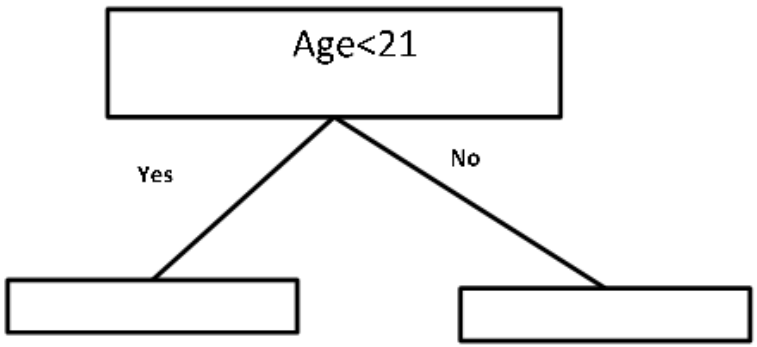
Considering 'Place of living=Village' as the separator

| Average PR2 score of Left side instances | Average PR2 score of Right side instances | Sum of square residuals (SSR) | SSR/2 |
|---|---|---|---|
| -0.633333334 | 0.63333333 | 9.786666667 | 4.893333333 |

Minimum SSR/2= 4.89333333

**Minimum SSR/2 for Age=3.67**

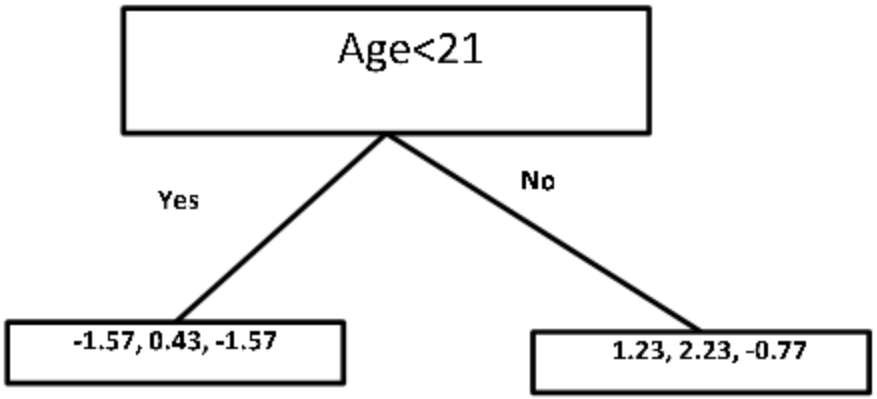
**Minimum SSR/2 Place of Living=4.89**

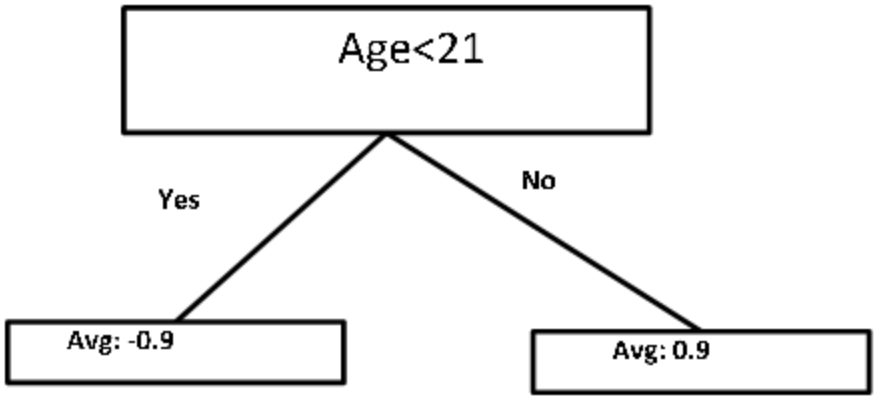


# Gradient Boosting for Regression Example

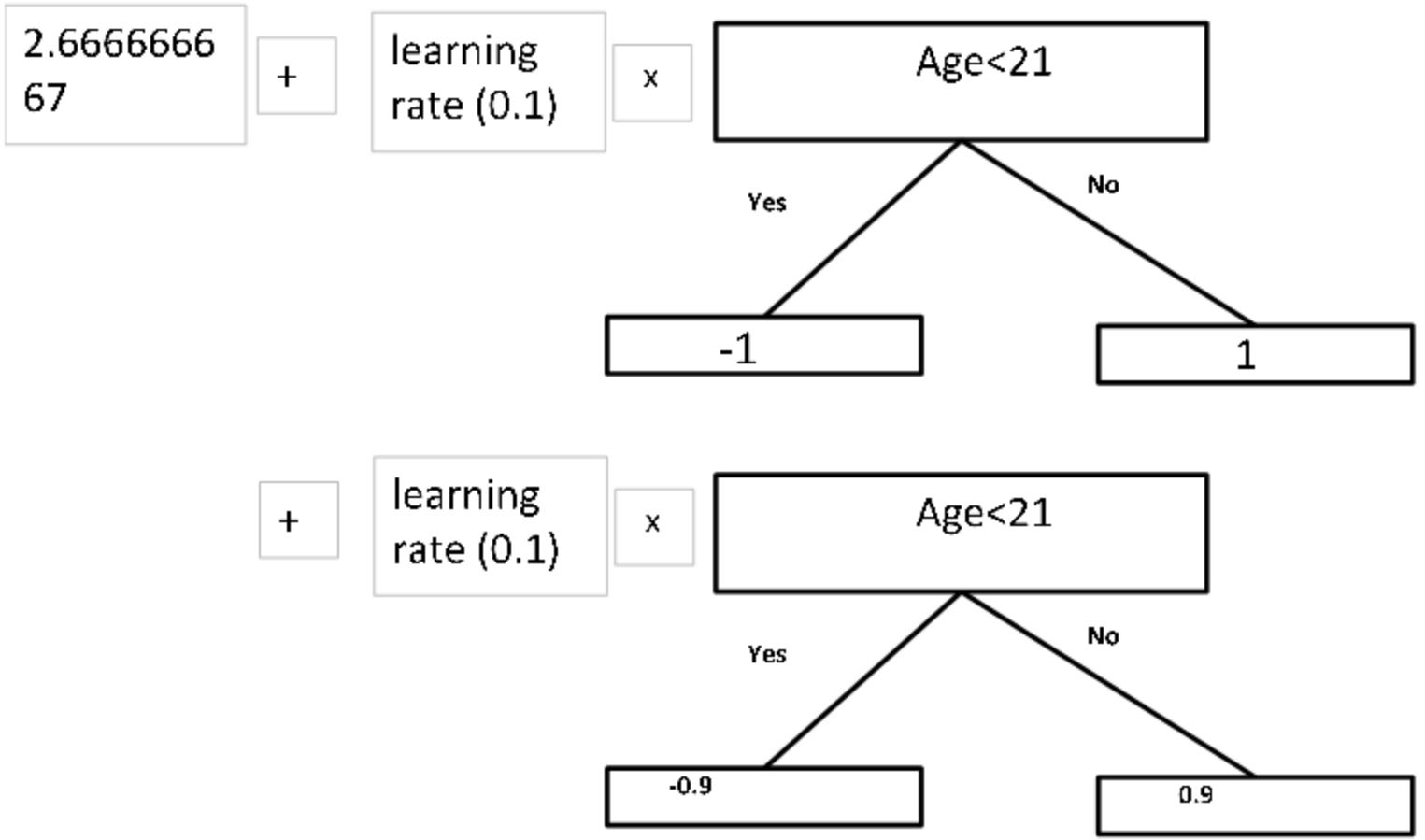
| | Age | Pseudo Residual 2 (PR2) | | Average PR2 score of Left side instances | Average PR2 score of Right side instances | Sum of square residuals (SSR)/2 |
|---|---|---|---|---|---|---|
| 1 | 8 | -1.566666667 | Considering age 9 as the separator | -1.56666667 | 0.31333333 | 4.624 |
| 2 | 10 | 0.433333333 | Considering age 11 as the separator | -0.56666667 | 0.28333333 | 5.615 |
| 3 | 12 | -1.566666667 | Considering age 21 as the separator | -0.9 | 0.9 | 3.666666667 |
| 4 | 30 | 1.233333333 | Considering age 32.5 as the separator | -0.36666667 | 0.73333333 | 5.29 |
| 5 | 35 | 2.233333333 | Considering age 47.5 as the separator | 0.153333333 | -0.76666667 | 5.744 |
| 6 | 60 | -0.766666667 | | | | |

Minimum SSR/2= 3.66666667



# Gradient Boosting for Regression Example



# Gradient Boosting for Regression Example

| Observations | Age | Place of living | IQ-score | Average IQ Score | Pseudo Residual 1 | Predicted IQ Score | Pseudo Residual 2 | Predicted IQ Score | Pseudo Residual 3 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 8 | Village | 1 | 2.666667 | -1.66667 | 2.5666667 | -1.566666667 | 2.4766667 | -1.476667 |
| 2 | 10 | City | 3 | | 0.33333 | 2.5666667 | 0.433333333 | 2.4766667 | 0.5233333 |
| 3 | 12 | Village | 1 | | -1.66667 | 2.5666667 | -1.566666667 | 2.4766667 | -1.476667 |
| 4 | 30 | Village | 4 | | 1.33333 | 2.7666667 | 1.233333333 | 2.6566667 | 1.3433333 |
| 5 | 35 | City | 5 | | 2.33333 | 2.7666667 | 2.233333333 | 2.6566667 | 2.3433333 |
| 6 | 60 | City | 2 | | -0.66667 | 2.7666667 | -0.766666667 | 2.6566667 | -0.656667 |

| Observations | Age | Place of living | Pseudo Residual 3 |
|---|---|---|---|
| 1 | 8 | Village | -1.476666667 |
| 2 | 10 | City | 0.523333333 |
| 3 | 12 | Village | -1.476666667 |
| 4 | 30 | Village | 1.343333333 |
| 5 | 35 | City | 2.343333333 |
| 6 | 60 | City | -0.656666667 |

# Gradient Boosting for Classification

| Observations | Age | Place of living | IQ-score | IQ-score |
|---|---|---|---|---|
| 1 | 8 | Village | 1 | Low |
| 2 | 10 | City | 3 | Low |
| 3 | 12 | Village | 1 | Low |
| 4 | 30 | Village | 4 | High |
| 5 | 35 | City | 5 | High |
| 6 | 60 | City | 2 | Low |

| Observations | Age | Place of living | IQ-score |
|---|---|---|---|
| 1 | 8 | Village | Low |
| 2 | 10 | City | Low |
| 3 | 12 | Village | Low |
| 4 | 30 | Village | High |
| 5 | 35 | City | High |
| 6 | 60 | City | Low |

| Observations | Age | Place of living | IQ-score | Encoded IQ-score |
|---|---|---|---|---|
| 1 | 8 | Village | Low | 1 |
| 2 | 10 | City | Low | 1 |
| 3 | 12 | Village | Low | 1 |
| 4 | 30 | Village | High | 0 |
| 5 | 35 | City | High | 0 |
| 6 | 60 | City | Low | 1 |

# Gradient Boosting for Classification

| Observations | Age | Place of living | IQ-score | Encoded IQ-score (Y) | *Predicted probability* | Residual |
|---|---|---|---|---|---|---|
| 1 | 8 | Village | **Low** | 1 | 0.166666667 | 0.833333333 |
| 2 | 10 | City | **Low** | 1 | 0.166666667 | 0.833333333 |
| 3 | 12 | Village | **Low** | 1 | 0.166666667 | 0.833333333 |
| 4 | 30 | Village | **High** | 0 | 0.166666667 | -0.166666667 |
| 5 | 35 | City | **High** | 0 | 0.166666667 | -0.166666667 |
| 6 | 60 | City | **Low** | 1 | 0.166666667 | 0.833333333 |

# Gradient Boosting for Classification

| | |
|---|---|
| F_Low= | 4 |
| F_High= | 2 |
| ln(Odds)=ln(F_Low/F_High)= | 0.69314718 |
| P_Low= | 0.66666667 |

P_Low > 0.5 So, Consider all of the instances have Low IQ score i.e 1 (Low)

$$\text{Probability} = \frac{e^{\ln(Odds)}}{1+e^{\ln(Odds)}}$$

| Observations | Age | Place of living | IQ-score | Encoded IQ-score (Y) | *Predicted probability* | Residual | log(Odds) |
|---|---|---|---|---|---|---|---|
| 1 | 8 | Village | Low | 1 | 0.666666667 | 0.333333333 | 0.693147181 |
| 2 | 10 | City | Low | 1 | 0.666666667 | 0.333333333 | 0.693147181 |
| 3 | 12 | Village | Low | 1 | 0.666666667 | 0.333333333 | 0.693147181 |
| 4 | 30 | Village | High | 0 | 0.666666667 | -0.666666667 | 0.693147181 |
| 5 | 35 | City | High | 0 | 0.666666667 | -0.666666667 | 0.693147181 |
| 6 | 60 | City | Low | 1 | 0.666666667 | 0.333333333 | 0.693147181 |

# Gradient Boosting for Classification

| Observations | Age | Place of living | Pseudo Residual |
|---|---|---|---|
| 1 | 8 | Village | 0.333333333 |
| 2 | 10 | City | 0.333333333 |
| 3 | 12 | Village | 0.333333333 |
| 4 | 30 | Village | -0.666666667 |
| 5 | 35 | City | -0.666666667 |
| 6 | 60 | City | 0.333333333 |

| | Age | Pseudo Residual (PR) | | Average PR score of Left side instances | Average PR score of Right side instances | Sum of square residuals (SSR)/2 |
|---|---|---|---|---|---|---|
| 1 | 8 | 0.333333333 | Considering age 9 as the separator | 0.333333333 | -0.066666667 | 0.6 |
| 2 | 10 | 0.333333333 | Considering age 11 as the separator | 0.333333333 | -0.166666667 | 0.5 |
| 3 | 12 | 0.333333333 | Considering age 21 as the separator | 0.333333333 | -0.333333333 | 0.333333333 |
| 4 | 30 | -0.666666667 | Considering age 32.5 as the separator | 0.083333333 | -0.166666667 | 0.625 |
| 5 | 35 | -0.666666667 | Considering age 47.5 as the separator | -0.066666667 | 0.333333333 | 0.6 |
| 6 | 60 | 0.333333333 | | | | |

**Minimum SSR/2=** 0.333333333

# Gradient Boosting for Classification

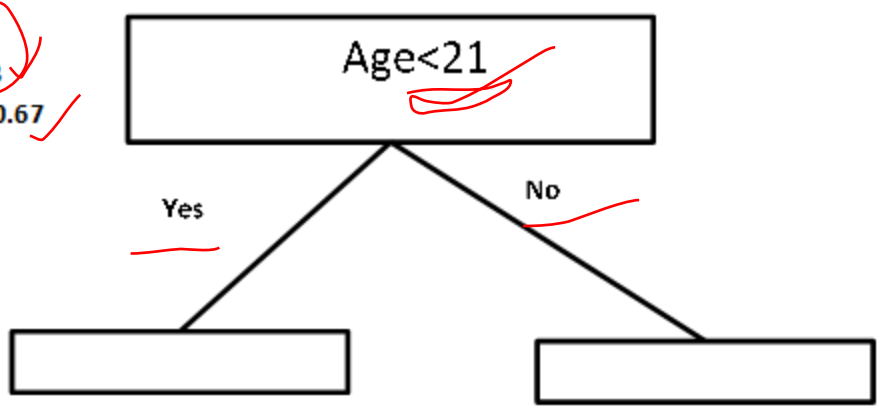
| Place of living | Encoded Place of living | Pseudo Residual (PR) |
|---|---|---|
| Village | 1 | 0.333333 |
| City | 2 | 0.333333 |
| Village | 1 | 0.333333 |
| Village | 1 | -0.666667 |
| City | 2 | -0.666667 |
| City | 2 | 0.333333 |

Considering 'Place of living=Village' as the separator

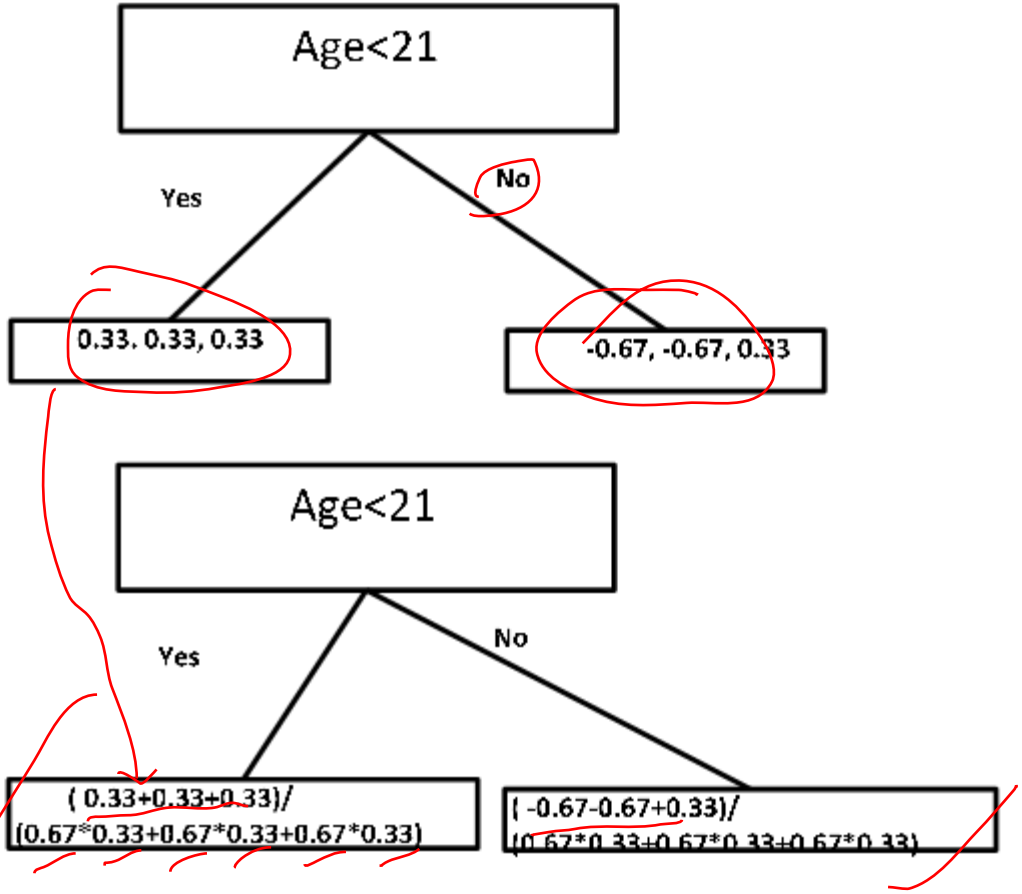
| Average PR score of Left side instances | Average PR score of Right side instances | Sum of square residuals (SSR) | SSR/2 |
|---|---|---|---|
| 3.70074E-17 | 3.70074E-17 | 1.333333333 | 0.666666667 |

Minimum SSR/2= 0.666666667

Minimum SSR/2 for Age=0.33
Minimum SSR/2 Place of Living=0.67

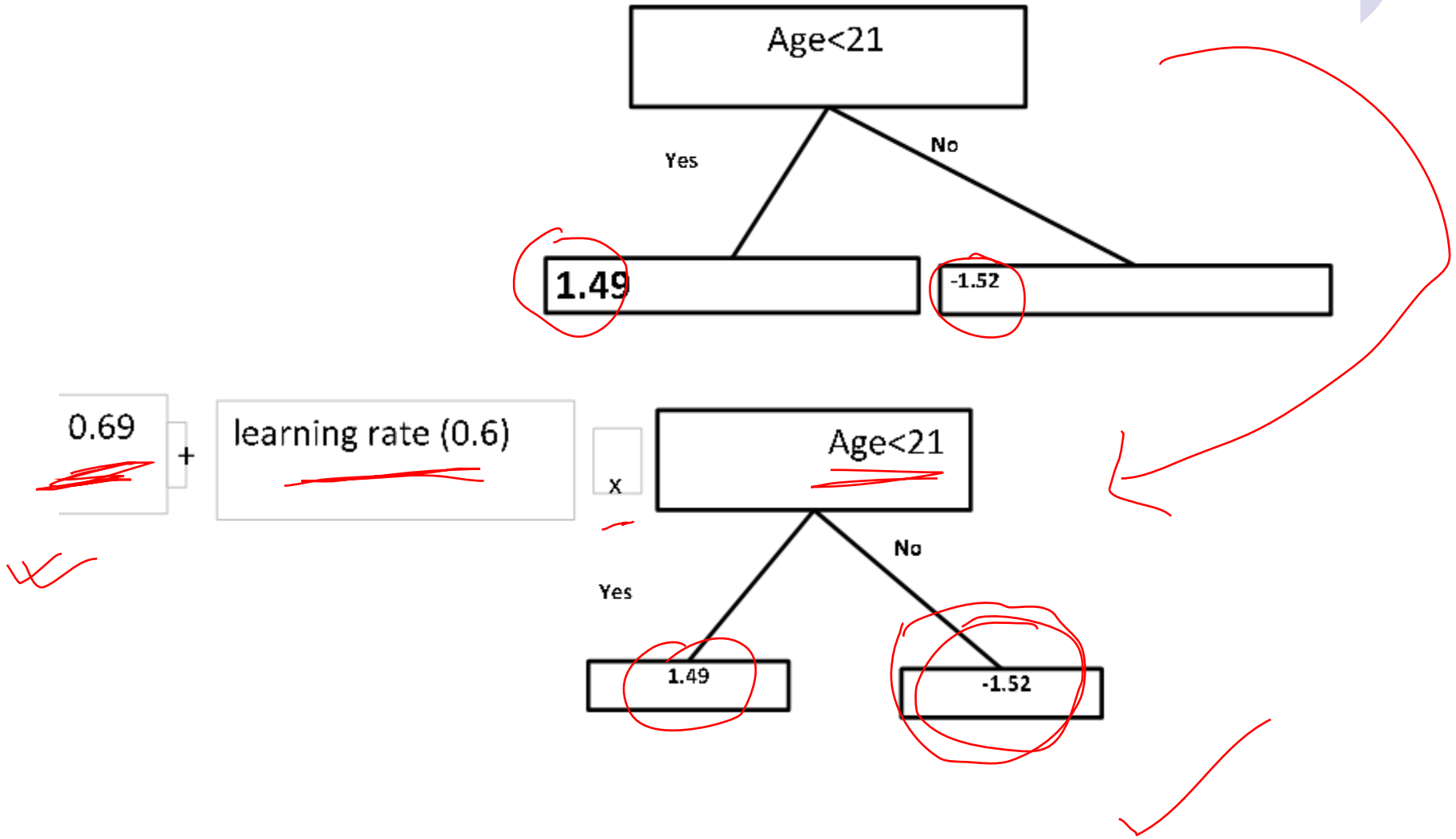


# Gradient Boosting for Classification




Sum of Residual/(Sum of (the Previous Probability * (1-the Previous Probability))

# Gradient Boosting for Classification



# Gradient Boosting for Classification

| Observations | Age | Place of living | IQ-score | Encoded IQ-score (Y) | *Predicted probability* | Residual | log(Odds) |
|---|---|---|---|---|---|---|---|
| 1 | 8 | Village | **Low** | 1 | 0.82977027 | 0.17022973 | 1.584 |
| 2 | 10 | City | **Low** | 1 | 0.82977027 | 0.17022973 | 1.584 |
| 3 | 12 | Village | **Low** | 1 | 0.82977027 | 0.17022973 | 1.584 |
| 4 | 30 | Village | **High** | 0 | 0.444726821 | -0.444726821 | -0.222 |
| 5 | 35 | City | **High** | 0 | 0.444726821 | -0.444726821 | -0.222 |
| 6 | 60 | City | **Low** | 1 | 0.444726821 | 0.555273179 | -0.222 |

# References


- 1. Boosting : foundations and algorithms by Robert E. Schapire and Yoav Freund.
- **MIT press**
- https://mitpress.mit.edu/sites/default/files/titles/content/boosting_foundations_algorithms/toc.html#indx-1

- **2. Boosting and AdaBoost for Machine Learning**
- https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/

- **3. Quick Introduction to Boosting Algorithms in Machine Learning**
- https://www.analyticsvidhya.com/blog/2015/11/quick-introduction-boosting-algorithms-machine-learning/

- **4. github repo and kaggle kernel link for GBM from scratch:**
  https://www.kaggle.com/grroverpr/gradient-boosting-simplified/
  https://nbviewer.jupyter.org/github/groverpr/Machine-Learning/blob/master/notebooks/01_Gradient_Boosting_Scratch.ipynb
- 5. A detailed and intuitive explanation of gradient boosting: **How to explain gradient boosting** by Terence Parr and Jeremy Howard
- **6. Fast.ai github repo link** for DecisionTree from scratch **(Massive ML/DL related resources)**:
  https://github.com/fastai/fastai
- 7. Lecture by **Alexander Ihler** **in UCA**

  https://www.youtube.com/watch?v=sRktKszFmSk&t=311s

8. Machine Learning With Boosting: A Beginner's Guide by Scott Hartshorn, Kindle Edition

9. Machine Learning With Random Forests And Decision Trees: A Visual Guide For Beginners by Scott Hartshorn, Kindle Edition

**10. MIT open courseware (Lecture 17: Boosting)**

https://www.youtube.com/watch?v=UHBmv7qCey4

**11. Hands-On Machine Learning with R by** *Bradley Boehmke & Brandon Greenwell*

https://bradleyboehmke.github.io/HOML/