**Where Probability and Statistics used in ml:**

Let’s take ex of Iris where we plot a pair plot, in this we can clearly classify setossa from other two, but there are some minor overlapping in virginica and versicolor, now let’s say we’ve to classify a point **xq** since it’s between the virginica and versicolor, so instead of saying directly it as verginica or versicolr, we can say in terms of prob as:

It’s versicolor with 0.8 chance or 80% prob

And It’s virginica with 0.2 chance or 80% prob

And ofcourse it’s setossa with 0% prob.



**Variable**: A value that can be change at any time. It is not fixed.

**Random Variable** : It is a value from outcome where number of outcome is not fixed and value could be any real number

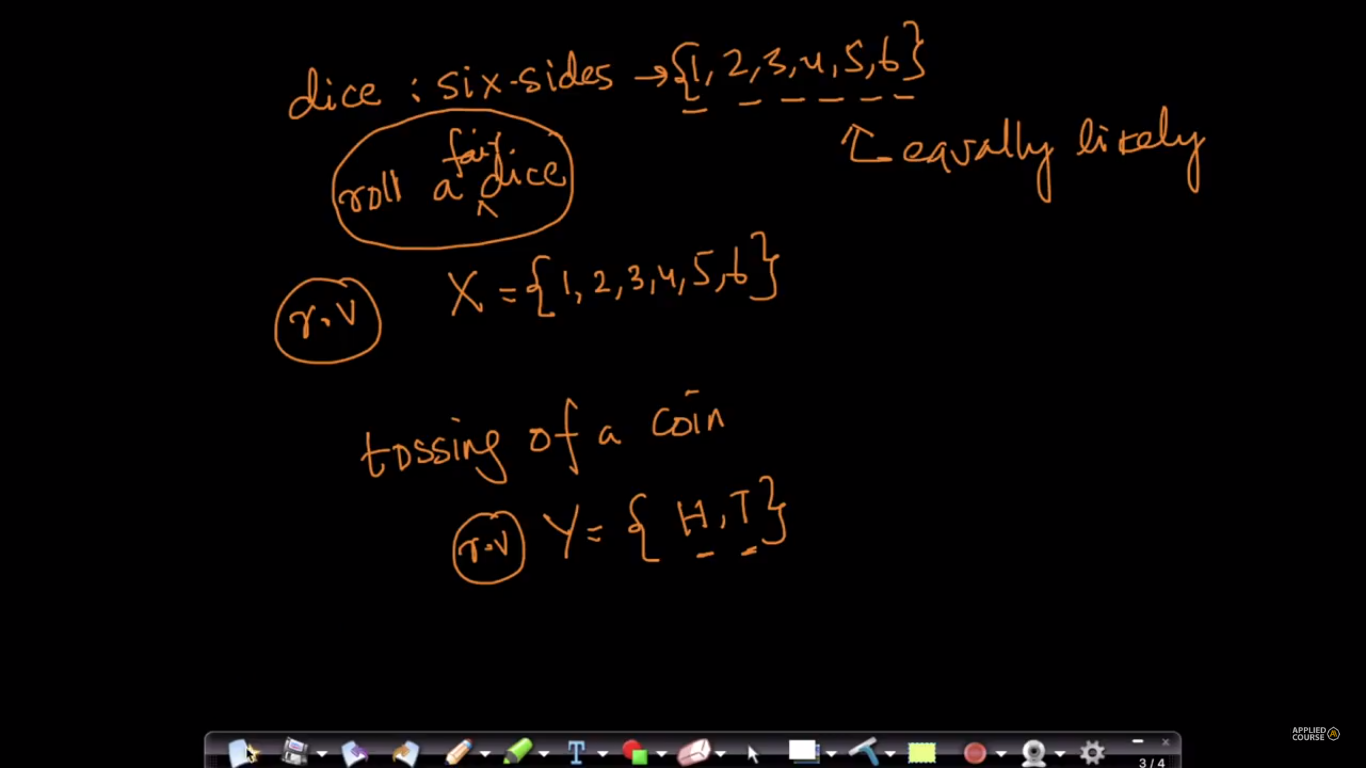
**Discrete random variable** : when outcome of random experiment is discrete i.e {5,6,4,7,8} it is discrete random variable

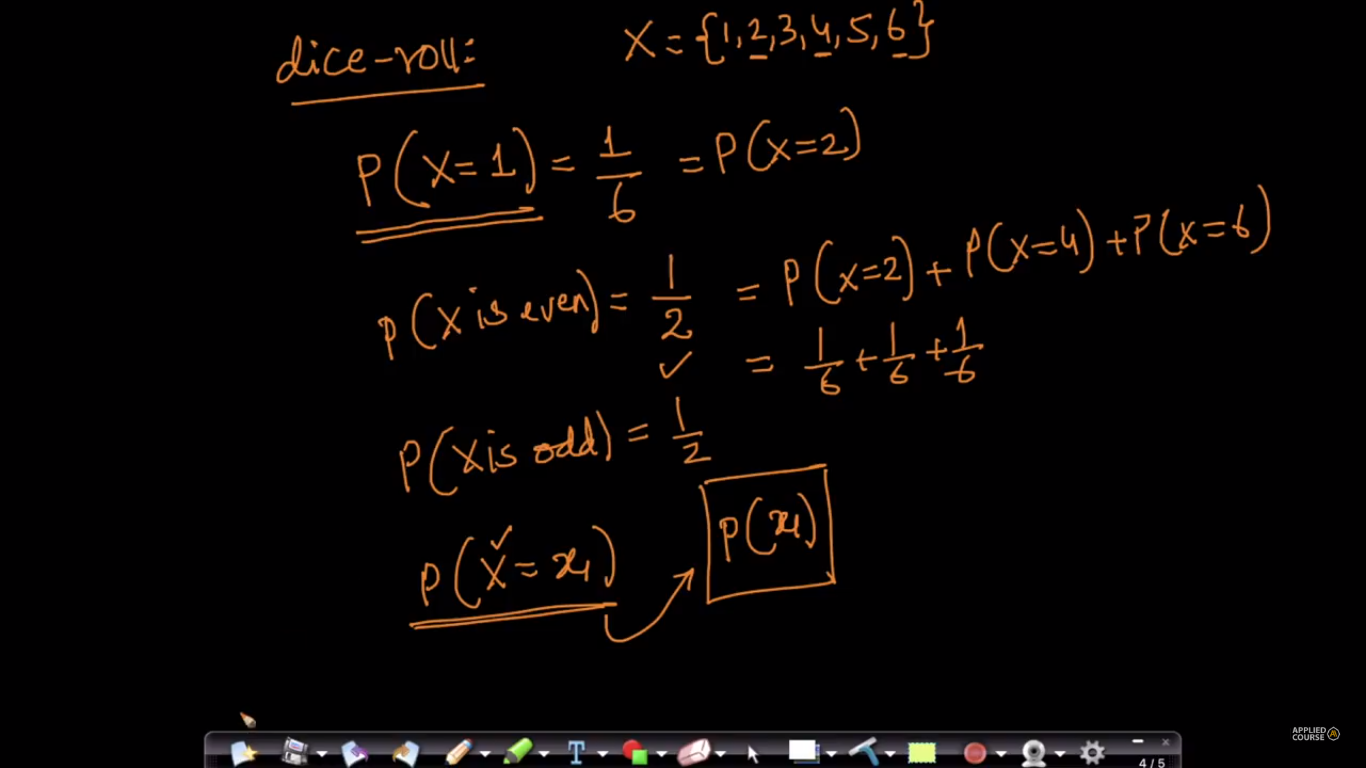
Ex: Suppose we flip a coin and count the number of heads. The number of heads could be any integer value between 0 and plus infinity. However, it could not be any number between 0 and plus infinity. We could not, for example, get 2.5 heads. Therefore, the number of heads must be a discrete variable.

**Continuous random variable**: when it is continuous i,e{4.5, 6.7, 8.2} it is continuous random variable or taking any random variable which we can't define at the starting of the experiment, it can be any value from –infinite to +infinite.

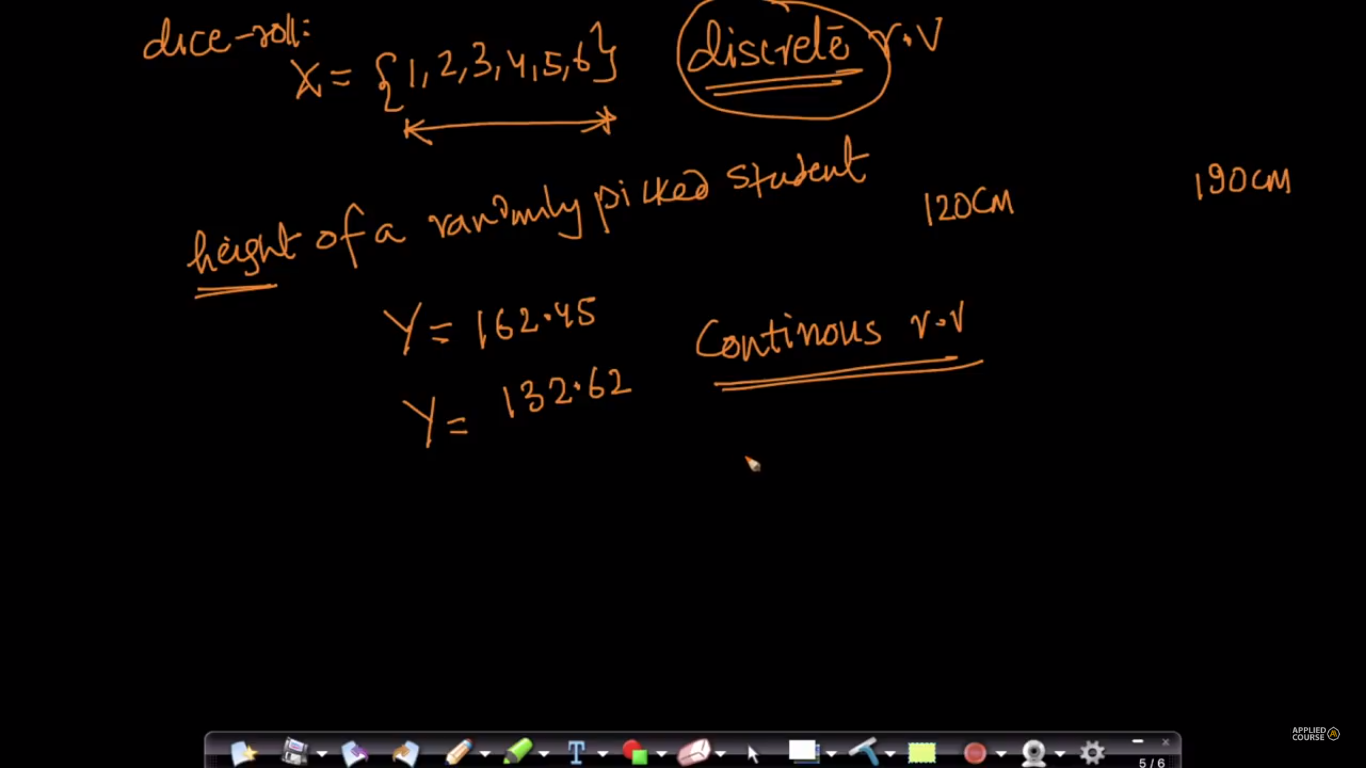
Ex: Suppose the fire department mandates that all fire fighters must weigh between 150 and 250 pounds. The weight of a fire fighter would be an example of a continuous variable; since a fire fighter's weight could take on any value between 150 and 250 pounds.

Ex of discrete random variable is given in below fig



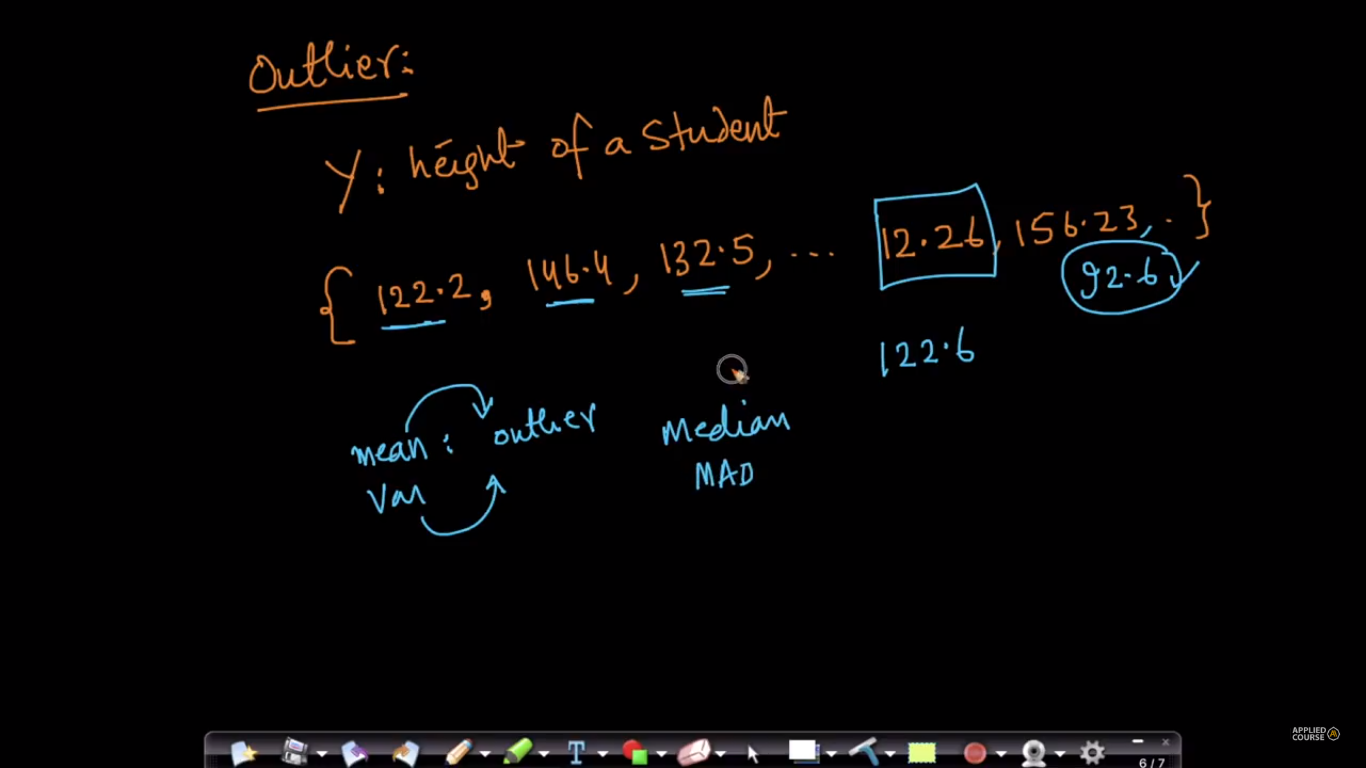


Ex of continuous random variable is given in below fig



In [regression analysis](https://www.stattrek.com/Help/Glossary.aspx?Target=Regression), a data point that diverges greatly from the overall pattern of data is called an outlier.

In more general usage, an outlier is an extreme value that differs greatly from other values in a [set](https://www.stattrek.com/Help/Glossary.aspx?Target=Set) of values. As a "rule of thumb", an extreme value is considered to be an outlier if it is at least 1.5 [interquartile ranges](https://www.stattrek.com/Help/Glossary.aspx?Target=Interquartile%20range)below the first [quartile](https://www.stattrek.com/Help/Glossary.aspx?Target=quartile)(Q1), or at least 1.5 interquartile ranges above the third quartile (Q3).



**For further reference:** <https://www.stattrek.com/probability-distributions/discrete-continuous.aspx>

For Any topic related to prob or stats.

<https://www.stattrek.com/statistics/dictionary.aspx>