Date - 20-May

* concept of underflow

→ representing decimal in computer using binary was always a challenge.
→ The problem is related to above
→ Underflow is condition when number is too small (near zero) ex → 0.000001, so it becomes problem to store and apply operation as the number may loose precision
→ It becomes problematic ex in case of biology (studying cells)
→ How underflow is related to Naive Bayes

Ex . cgpa iq placement

• new query {8.1, 81} → Y / No ?

Yes → P(Yes | 8.1, 81)

No → P(No | 8.1, 81)

$P(Y \mid 8.1, 81) = P(Y)\,(P(8.1 \mid Yes))\,(P(81 \mid Yes))$

This prob is related to → cgpa

This prob is relate to placement

- Hence, if there were 2500 columns, there will be 2500 probabilities. and each probability lies between $0 < x < 1$. and if they were to multiply the number will be very very small, and consider it as zero (underflow)
- Same will happen with P(No/ given), both probas will be zero due to underflow

Solution for underflow ⇒ log probabilities

- $\log(a,b) = \log a + \log b$

Hence · $\log(p(a)\,p(b)\,p(c)\cdots) = \log(p(a) + \log(p(b)) + \ldots$

- So even if we were to have 2500 columns we can apply and convert into log and after adding we get log of probabilities
- whoever log probability is more, that class will be assigned.