

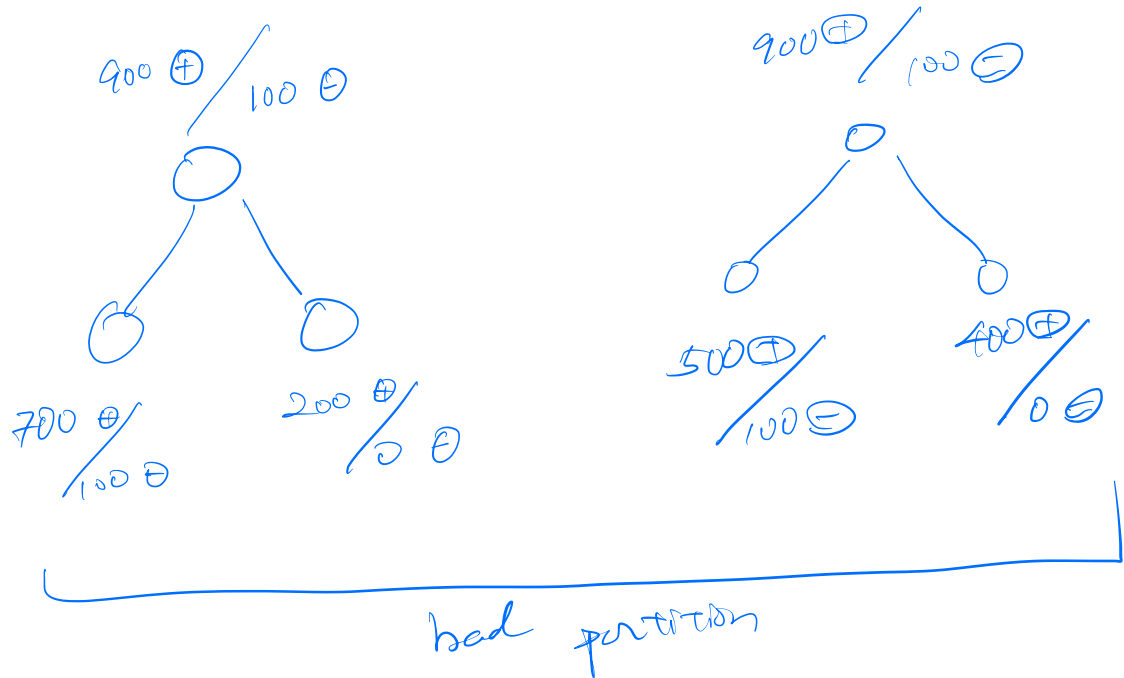
where can I ski?

⇒ cannot use a simple linear model for this kind of data...

→ decision tree

training not that slow (faster than neural network)

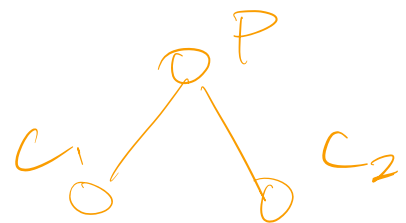
△ how do select features on node  
& value (page. 2)



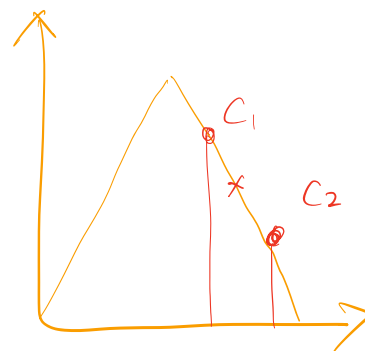
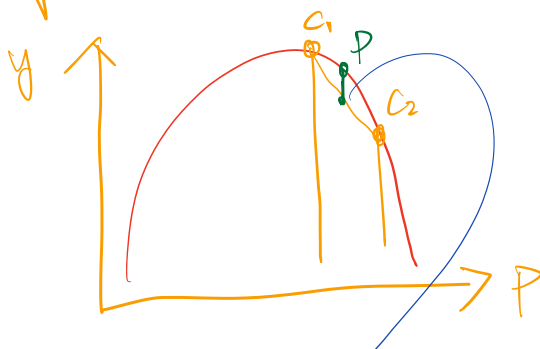
⇒ What kind of metric to be designed  
for better partition

maximum entropy

$$H(S) = - \sum_{C=1}^C p(C) \log p(C)$$



Why we use entropy?



1. Information gain  
 2. Gini  $1 - \sum p^2$

entropy low  
 } to judge whether the data is a good split or not

---

We can also use regression tree

---

$f_1, f_2, \dots, f_T$

$\sum_{i=1}^T w_i f_i$  : Bagging

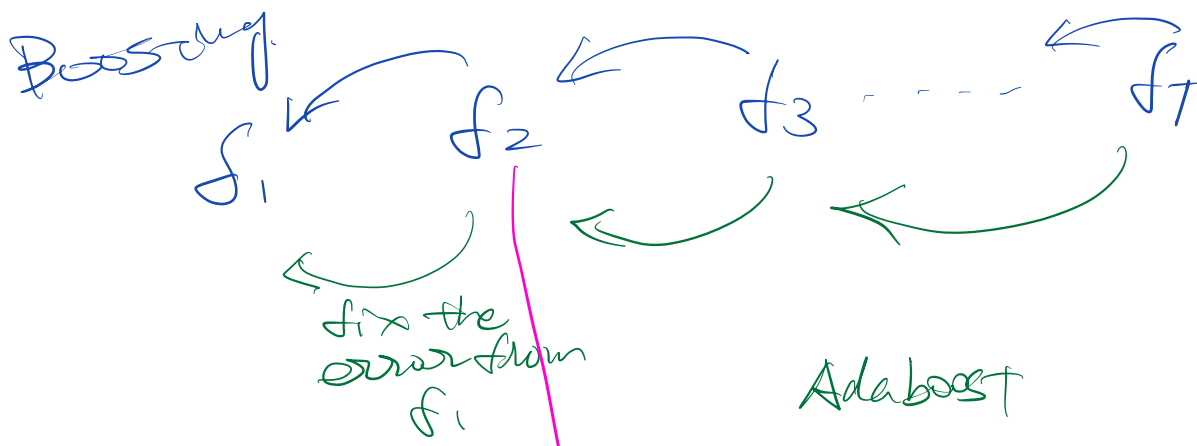
---

PIB: usually : give probability value.

---

Difference between Boosting & Bagging

---



↓ model to  
predict the error,  
cannot use independability.