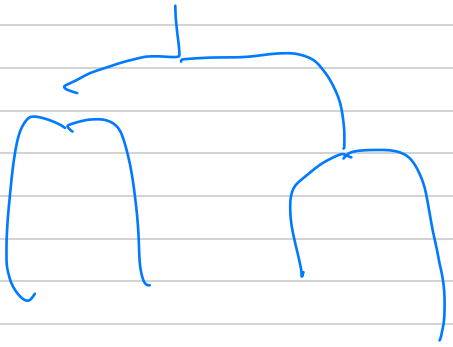
8-5-2025 **Regression Trees**

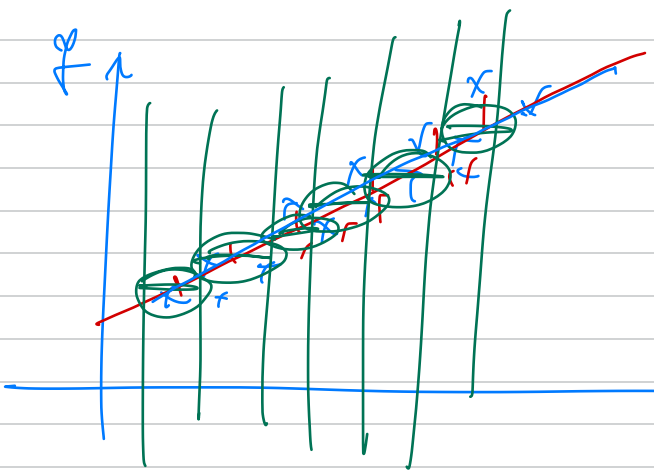
Decision Tree



Linear Regression

$$f(x) = b_0 + b_1 X_1 + \cdots + b_p X_p$$

1) $$f(x) = \sum_{i=}^{M} c_i \cdot 1(x \cdot \in R_i)$$



$y \; b_0 + b_1 X$.

RSS <<

2) Τα δέντρα έχουν πιο εύκολη εξήγηση της προβλεπτικής ανάλυσης

3) Αν υπάρχουν πολλές μεταβλητές $X_1, \ldots X_p$.

με regression : $Y = b_0 + b_1 X_1 + \cdots + b_k X_k \quad (k \leq p)$

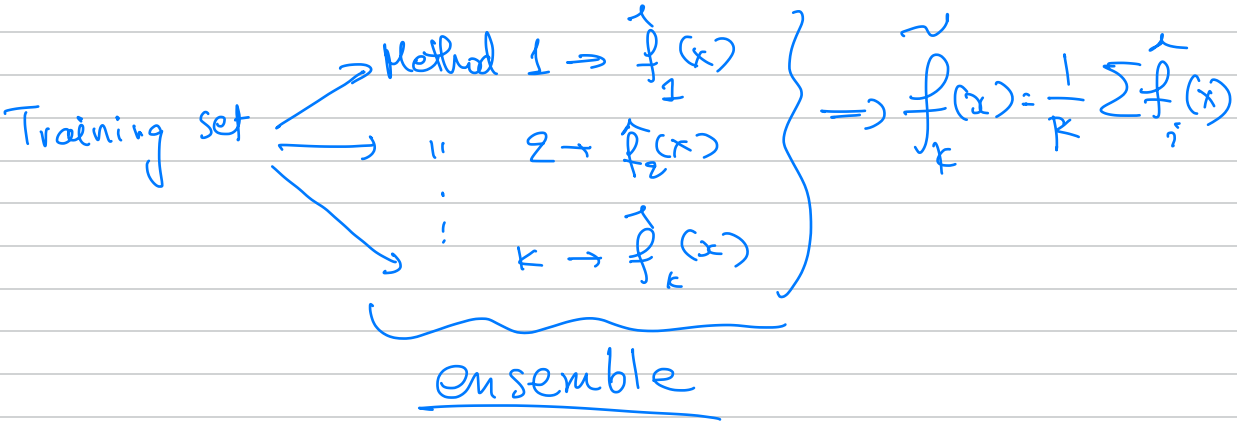
Ή τρόποι να αξιοποιηθεί η πληροφορία μιας μεταβλητής

Αυτό είναι διαφορετικό στο decision tree

4) Γενικά ένα "μεγάλο" decision tree έχει μεγάλο variance. //

Ensembles

Τεχνική ιδέα : model ensemble.

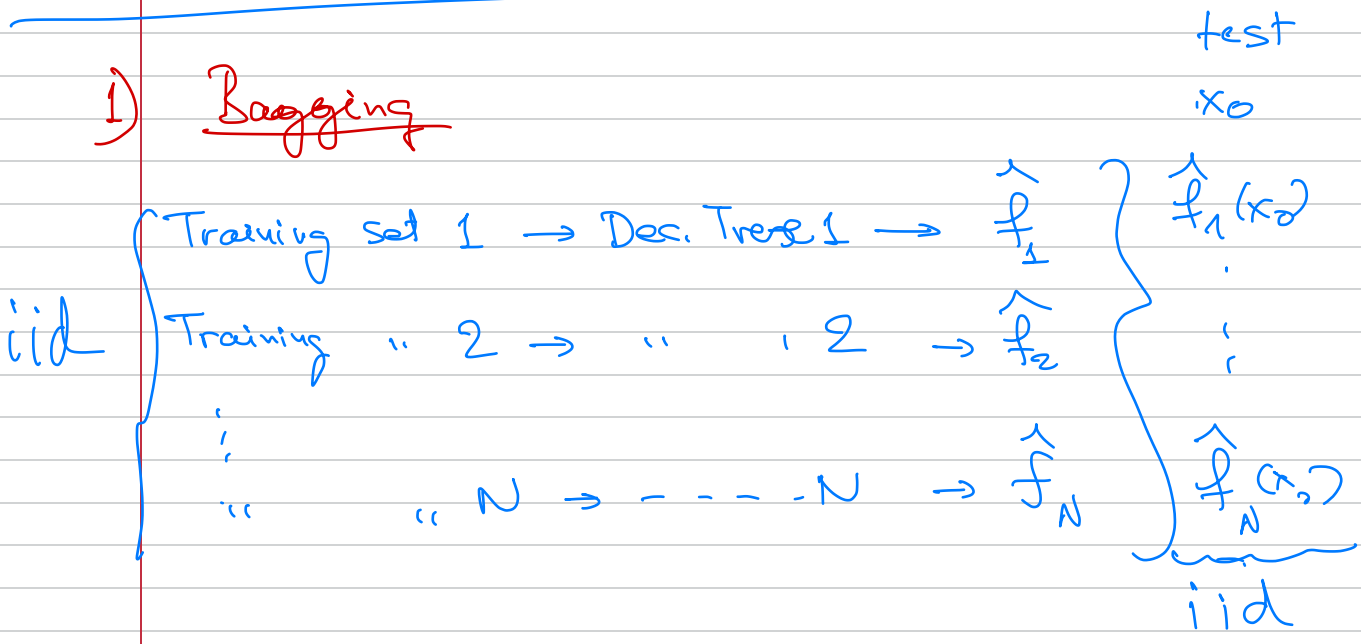


Η εφαρμογή εδώ

1) Bagging
2) Random Forests
3) Boosting

---

1) Bagging



Εστω $V(x_0) = \text{Var}\, \hat{f}_1(x_0)$

Αν $\bar{f}_N(x_0) = \frac{1}{N}\sum_{i=1}^{N} \hat{f}_i(x_0) \Rightarrow E\left[\bar{f}_N(x_0)\right] = E\left[\hat{f}_i(x_0)\right]$

$$Var\left[\bar{f}_N(x_0)\right] = \frac{1}{N} V(x_0)$$

Αν έχουμε μόνο ένα training set μεγέθους n ("τεράστιο n")

κ' πάρουμε N bootstrap δείγματα από αυτό

1) κάθε bootstrap set i.i.d από εμπειρική κατανομή του δείγματος

2) Η εμπειρική κατανομή του δείγματος ≈ πραγματική κατανομή των (X, Y) όταν n μεγάλο

Bagging Algorithm

1) Bootstrap samples $b = 1, \ldots, B$ από training set

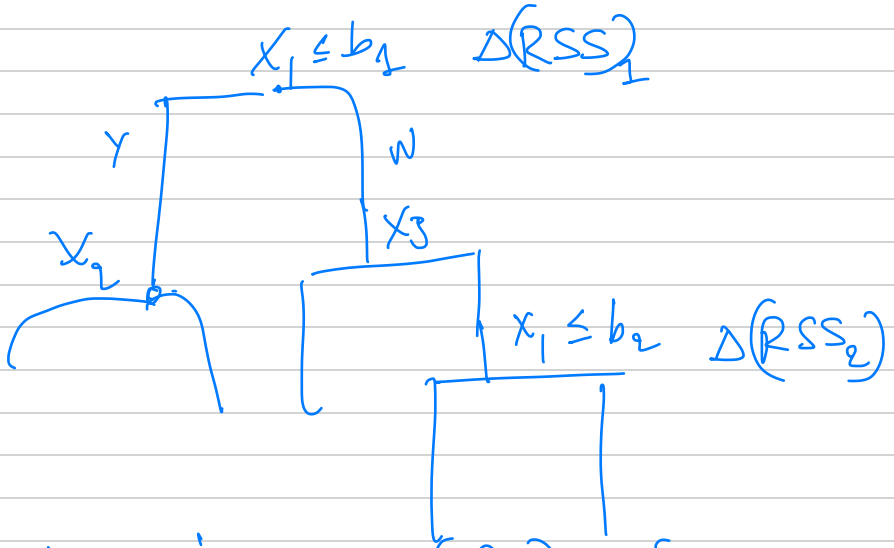
2) Dec. tree σε κάθε $b \rightarrow \hat{f}_b(x)$

3) $$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(x)$$

Variance importance factors (VIF) για μεταβλητή X_1

Σε κάθε δέντρο $b = 1, \ldots, B$

υπολογίζουμε τη συνολική μείωση του RSS που προκύπτει από διαίρεση με μεταβλητή X_1.



$(VIF)_{X_1}$: Συνολική μείωση $\Delta(RSS)$ εξ αιτίας της X_1 (μέσος όρος σε $b = 1, \ldots, B$)

2) Τυχαία Δέντρα

Έστω ότι η X_1 είναι πολύ σημαντική
Τότε θα υπάρχει στα περισσότερα bootstrap samples
και τα $\hat{f}_1, \ldots, \hat{f}_b$ θα έχουν
δεντρί συσχέτιση

Έστω X_1, X_2 i.d. αλλά $Cov(X_1, X_2) \neq 0$

$$Var\left(\frac{X_1+X_2}{2}\right) \neq \frac{1}{2} Var(X_1)$$

$$Var(X_1+X_2) = 2\,Var(X_1) + 2\,Cov(X_1, X_2)$$

$$Var\left(\frac{X_1+X_2}{2}\right) = \frac{1}{2} Var(X_1) + \frac{1}{2} Cov(X_1, X_2)$$