

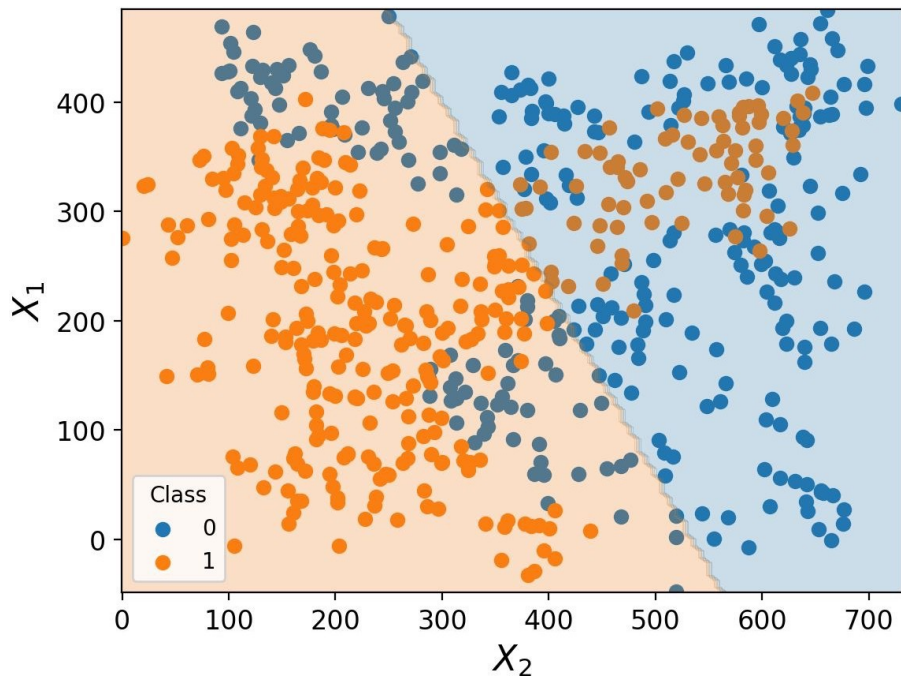
Ағаштар мен ансамбль үлгілері

Шешім ағаштары (Decision Tree)

Сызықтық классификацияның шектері

Сызықтық модельдер сызықтық шешім шекараларын жасайды

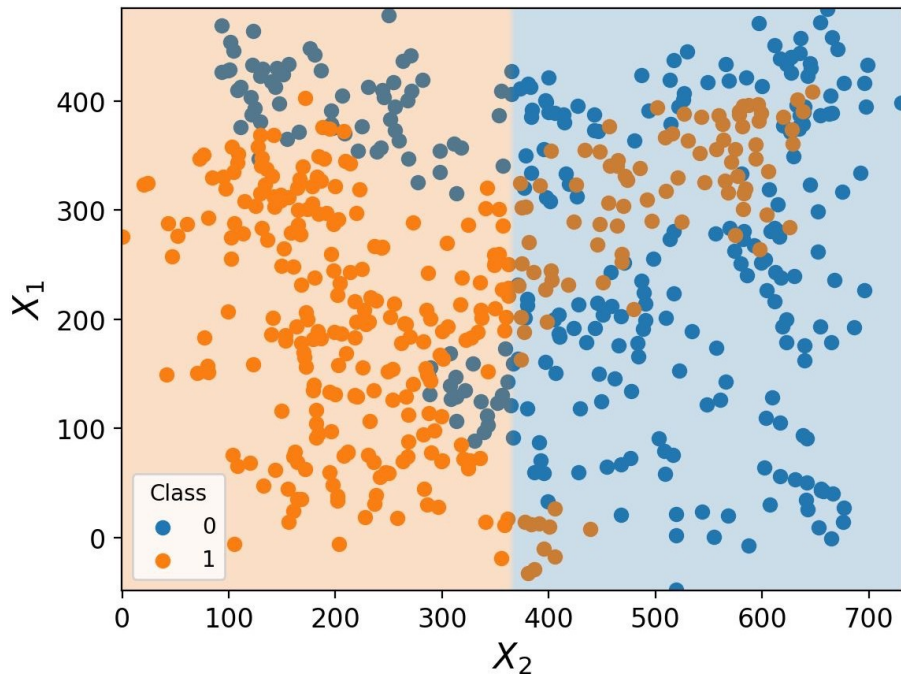
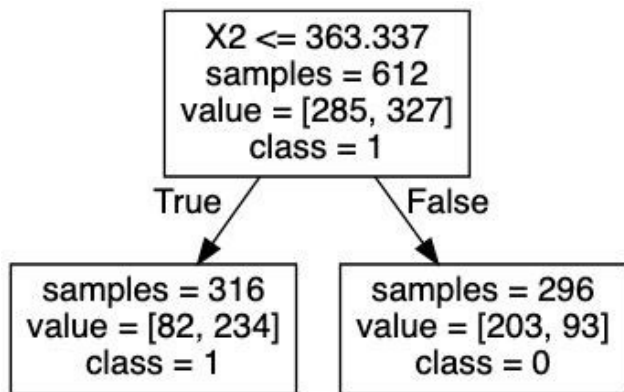
Бірақ олардың көпшілігін біріктірсек ше?



Шешім ағаштары

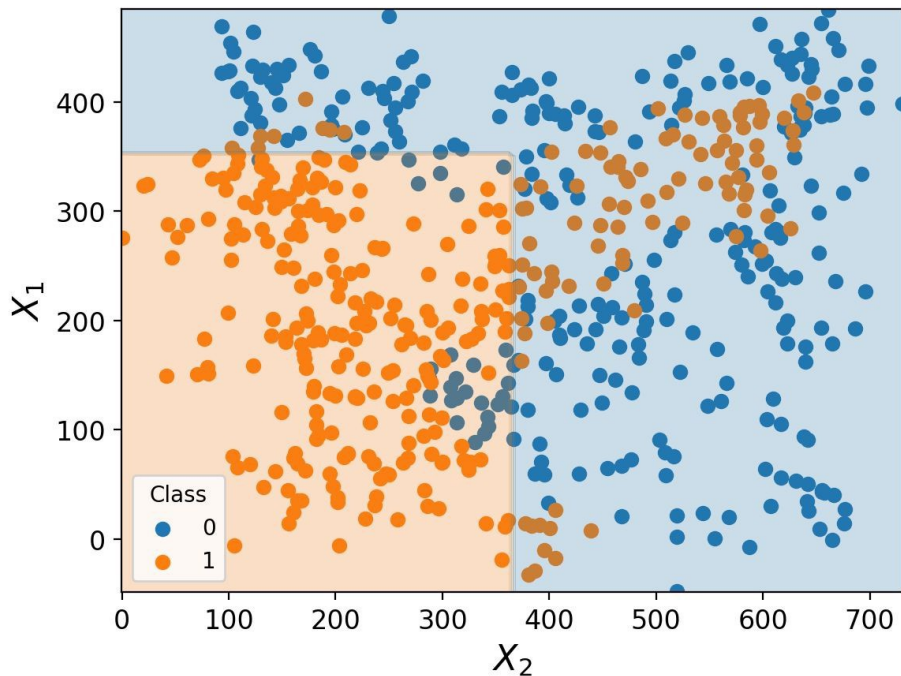
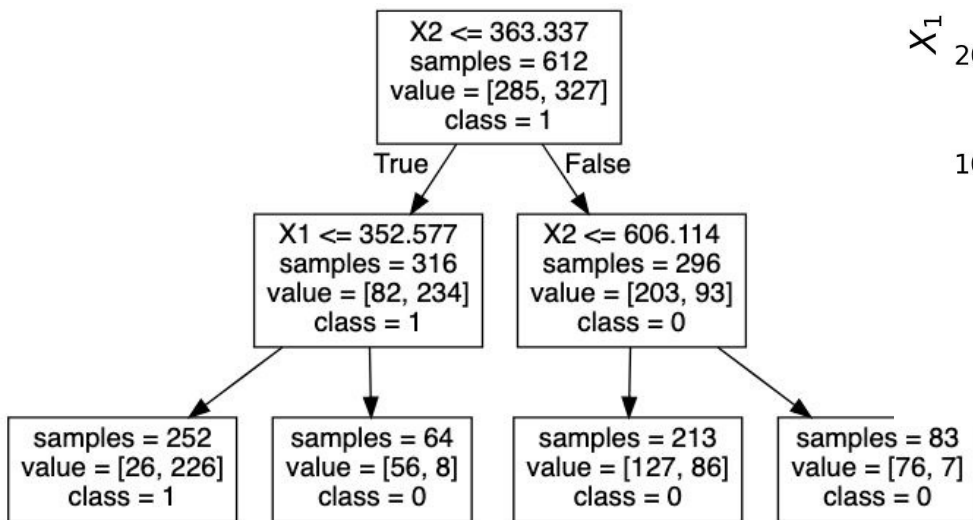
Өте қарапайым «ережеден»
бастаңыз:

Егер $X_2 \leq 363,337$,
содан кейін 1-сынып, басқасы
0-сынып



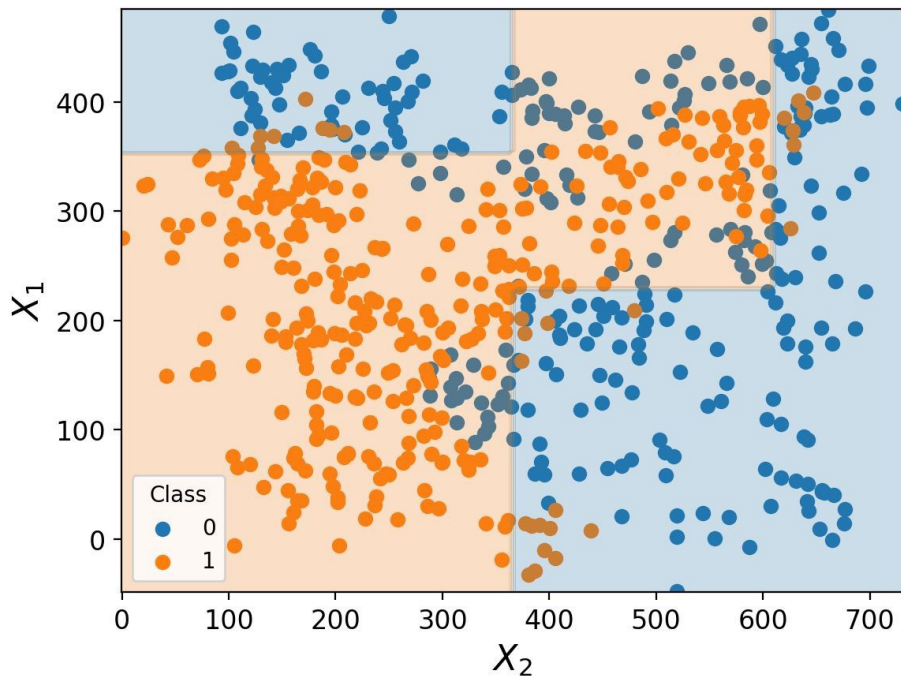
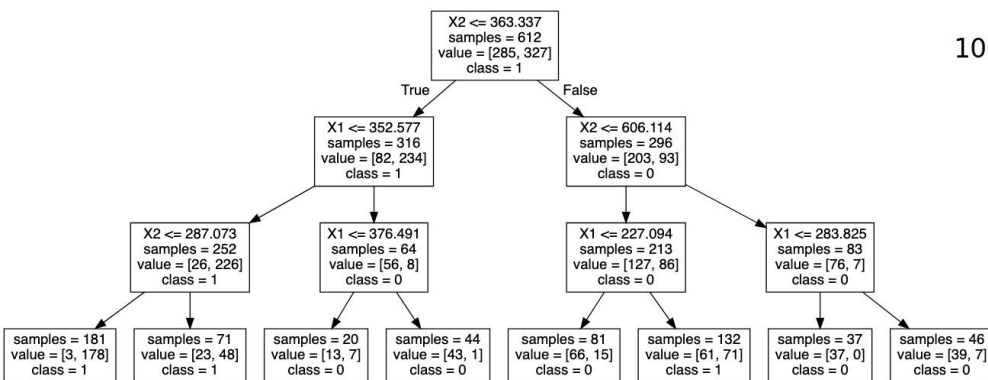
Шешім ағаштары

Қосымша if/else мәлімдемелерін қосыңыз, шешім шекарасына көбірек тараптар алыңыз.



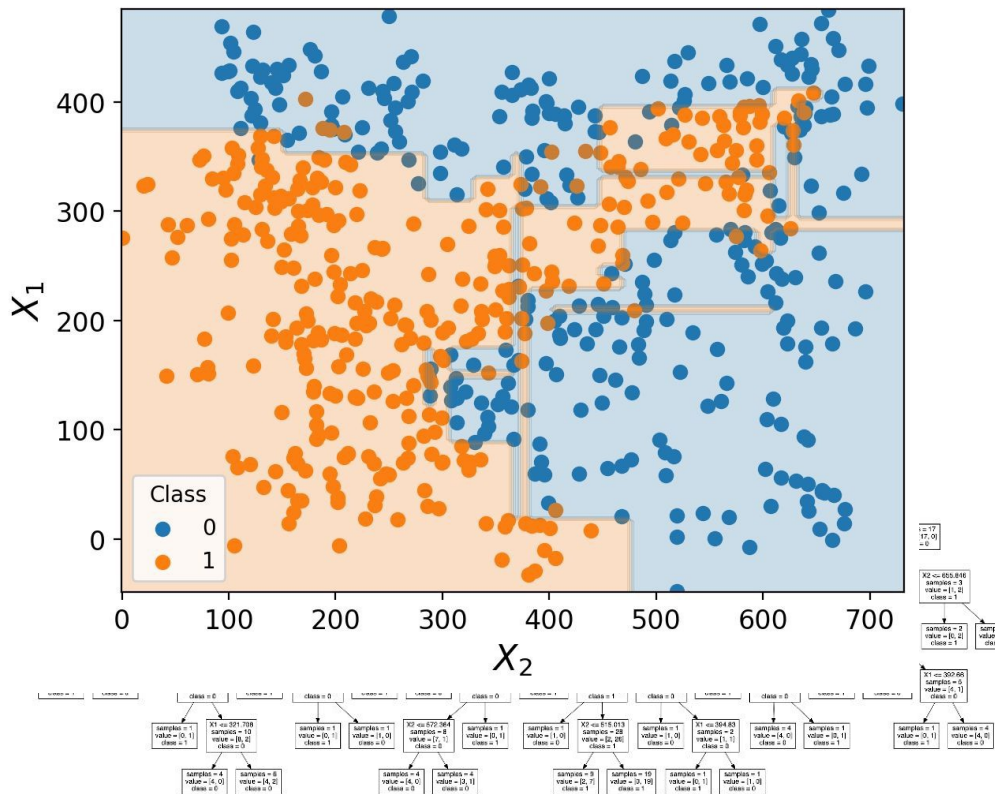
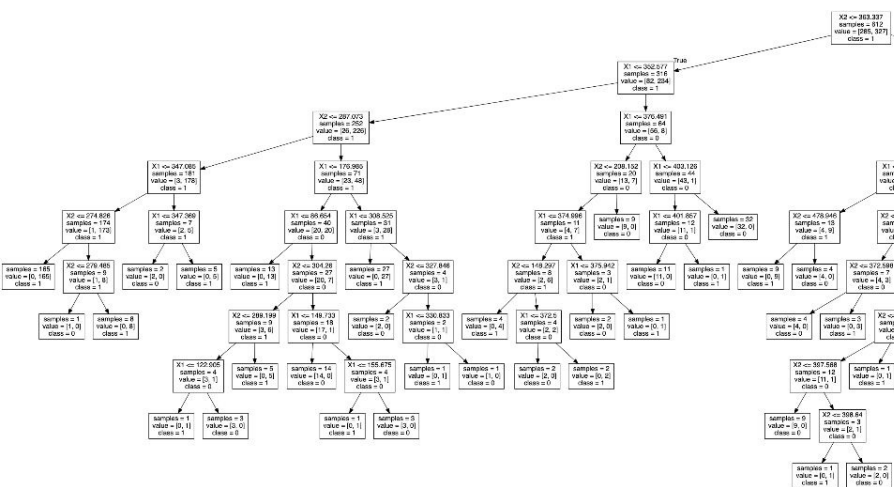
Шешім ағаштары

Қосымша if/else мәлімдемелерін қосыңыз, шешім шекарасына көбірек тараптар алыңыз.



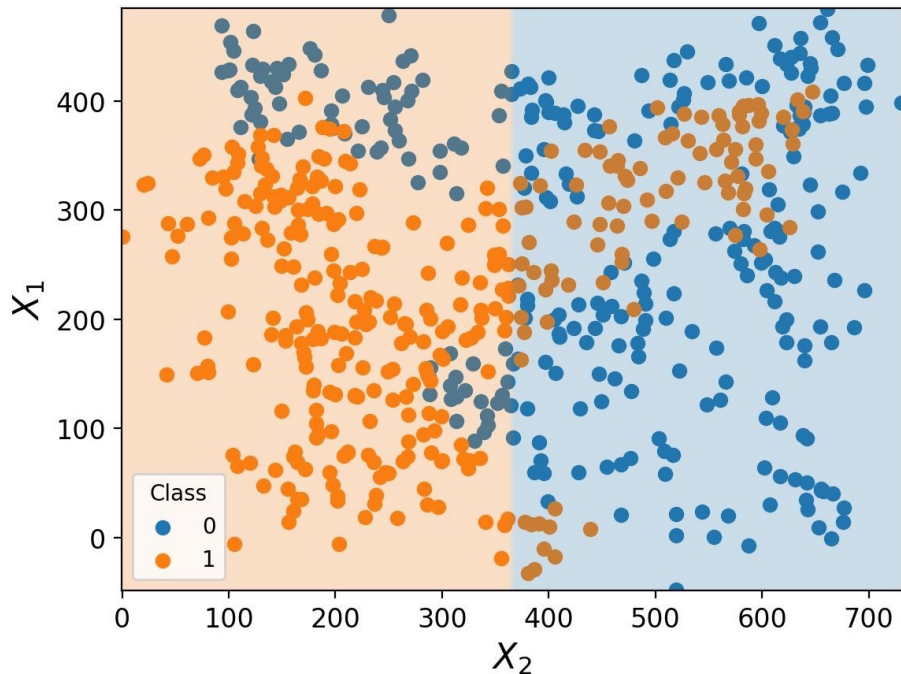
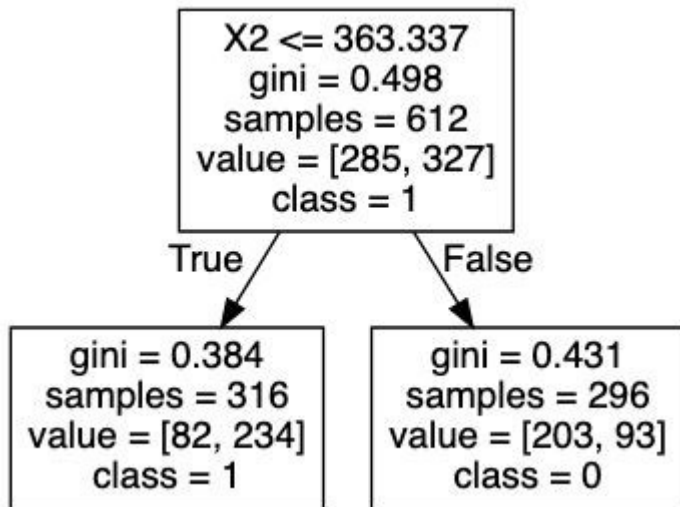
Шешім ағаштары

Қосымша if/else мәлімдемелерін қосыңыз, шешім шекарасына көбірек тараптар алыңыз.



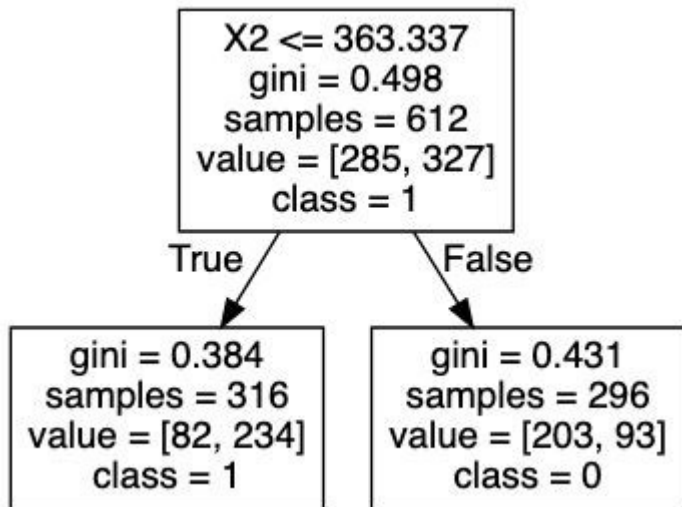
Шешім ағаштары - олар қалай өседі?

Әрбір түйінде мақсат сыныптарды барынша бөлетін мүмкіндік пен шекті табу болып табылады.



Шешім ағаштары - олар қалай өседі?

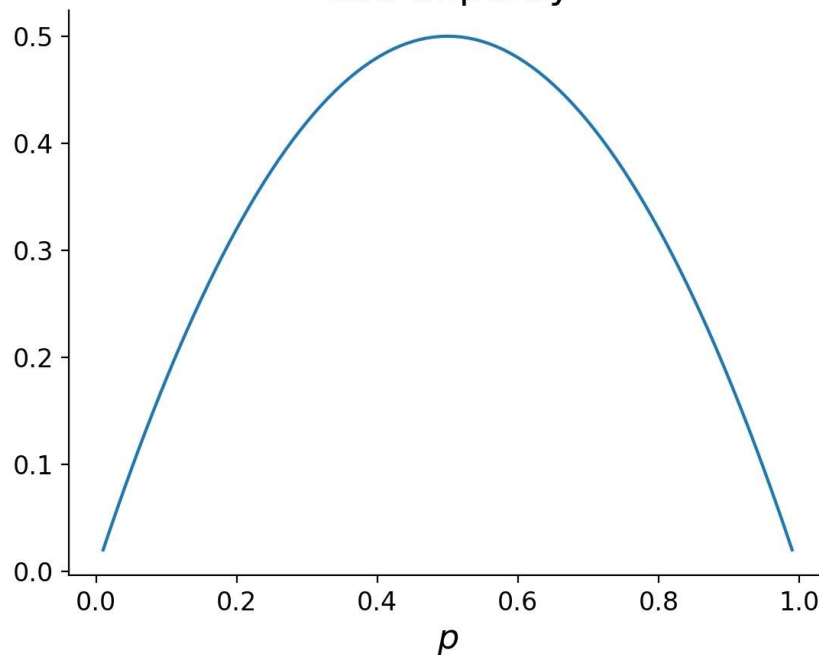
Джини қоспасы - бұл сыныптардың қаншалықты жақсы бөлінгенін көрсететін өлшем.



Түйіннен кейінгі оң
кластағы үлгілердің үлесі

$$Gini = 2p(1 - p)$$

Gini Impurity



Шешім ағаштары - олар қалай өседі?

Әр түрлі басқа да
«таза» шаралары

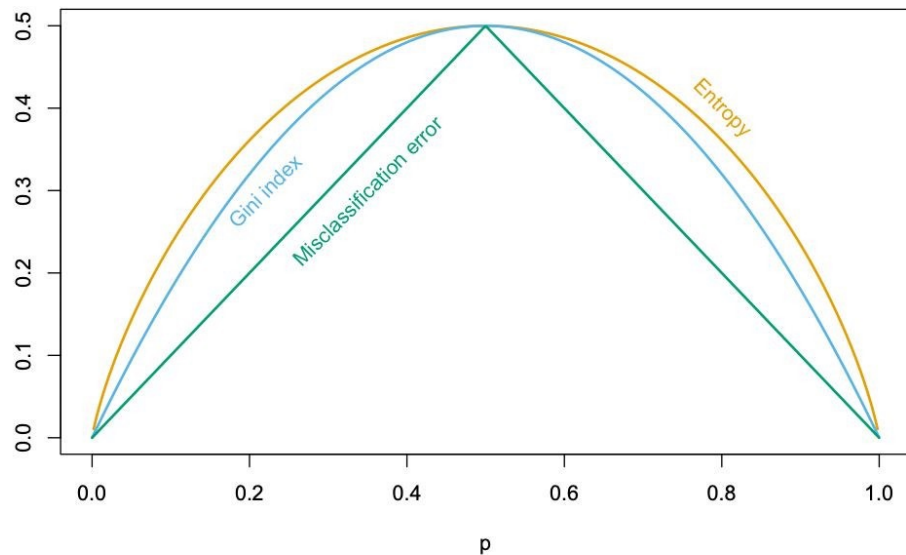
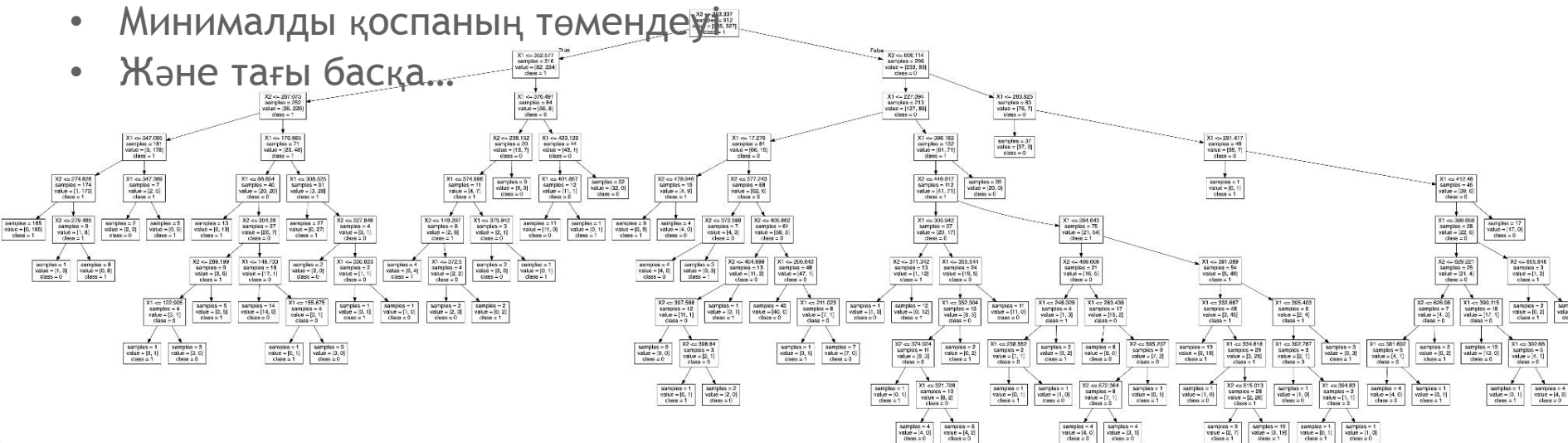


FIGURE 9.3. Node impurity measures for two-class classification, as a function of the proportion p in class 2. Cross-entropy has been scaled to pass through $(0.5, 0.5)$.

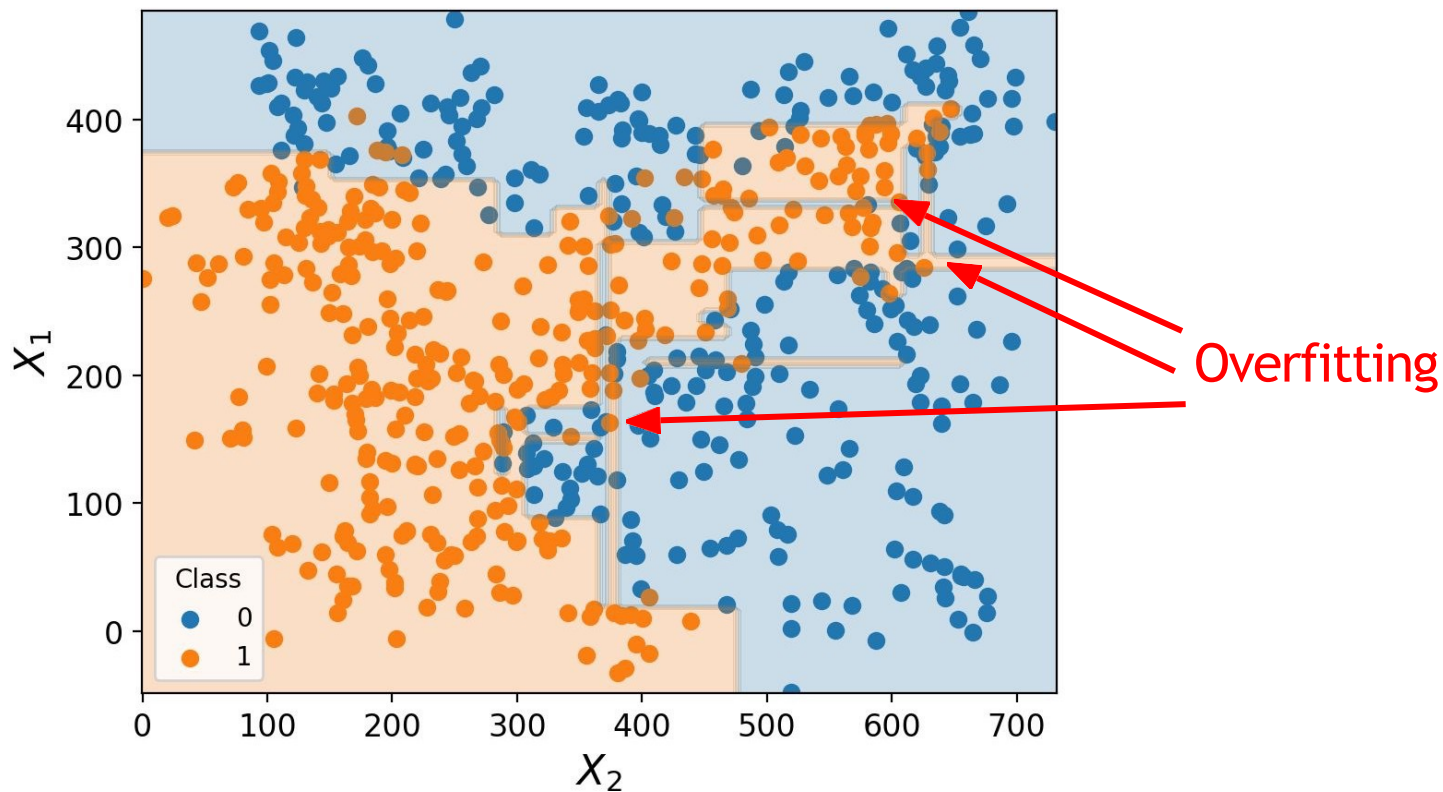
Шешім ағаштары - олар қалай өседі?

Кейбір кесу критерийлеріне дейін ағашты өсіруді жалғастырыңыз:

- Максималды тереңдік
- Жапырақ түйіндеріндегі ең аз үлгілерге жетті
- Минималды қоспаның төмендеуі
- Және тағы басқа...



Шешім ағаштарының шектері



Кездейсоқ ормандар (Random forester)

Кездейсоқ ормандар - ағаштар үшін орманды көру

Жалғыз шешім ағашының орнына көптеген ағаштар жасаңыз (орман!).

Бірақ кездейсоқтықты тудырыңыз. Әрбір ағаш үшін:

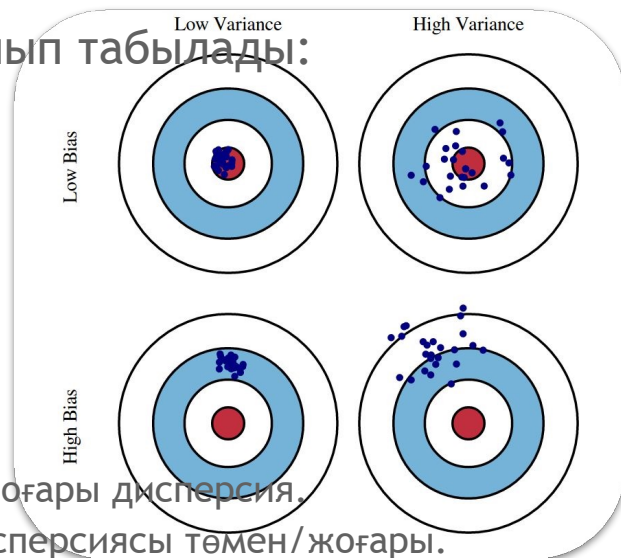
- Деректер жиынының жүктелу үлгісін жасаңыз (яғни ауыстыруы бар үлгі).
- Әрбір түйін үшін қандай мүмкіндікке бөлу керектігін шешкен кезде мүмкіндіктердің ішкі жиынын ғана қарастырыңыз.
- Әрбір сынып үшін болжау ұпайы үлгіні сол сыныпқа жіктейтін ағаштардың үлесі болып табылады.

Кездейсоқ ормандар - ағаштар үшін орманды көру

- Әрбір ағаш индукцияланған кездейсоқтыққа байланысты жалғыз шешім ағашы сияқты болжауда жақсы емес.
- Дегенмен, орман шамадан тыс тұрудың алдын алуға көмектеседі.
- Бұл шамадан тыс алдын-алу көбінесе әрбір ағаштың әлсіздігіне қарағанда күштірек, бұл жақсы жалпы үлгіге әкеледі.
- (Қолданудан тыс, бірақ бұл біржақтылық-дисперсиялық сауданың көрінісі).

Кездейсоқ ормандар - ағаштар үшін орманды көру

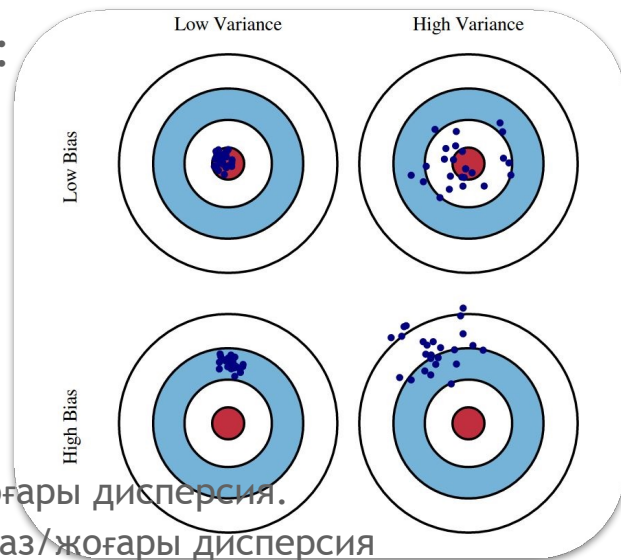
- (Қолданудан тыс, бірақ бұл біржақтылық-дисперсиялық сауданың көрінісі).
- Әрбір болжау қатесі мыналардың нәтижесі болып табылады:
 - Бағалау қатесі: мақсатты функция туралы болжамдарымызға байланысты.
 - Дисперсиялық қате: деректер жиынтығының ерекшеліктеріне байланысты.
 - Қайталанбайтын қате: бұл жерде ештеңе істей алмаймыз!
- Мысалдар :
 - Регрессияда төмен/жоғары ауытқу бар, бірақ төмен/жоғары дисперсия.
 - Шешім ағаштарының ауытқуы төмен/жоғары, бірақ дисперсиясы төмен/жоғары.
 - Кездейсоқ ормандарда төмен/жоғары ауытқу бар, бірақ төмен/жоғары дисперсия.



The typical bias/variance image
in all blog posts on the web!

Кездейсоқ ормандар - ағаштар үшін орманды көру

- (Қолданудан тыс, бірақ бұл біржақтылық-дисперсиялық сауданың көрінісі).
- Әрбір болжау қатесі нәтижесі болып табылады :
 - Бағалау қатесі: мақсатты функция туралы болжамдарымызға байланысты.
 - Дисперсиялық қате: деректер жиынтығының ерекшеліктеріне байланысты.
 - Қайталанбайтын қате: бұл жерде ештеңе істей алмаймыз!
- Мысалдар :
 - Регрессияда төмен/жоғары ауытқу бар, бірақ төмен/жоғары дисперсия.
 - Шешім ағаштарының төмен/жоғары ауытқуы бар, бірақ аз/жоғары дисперсия (шамадан тыс сәйкестік).
 - Кездейсоқ ормандар төмен/жоғары бейімділікке ие, бірақ аз/жоғары дисперсия (жалғыз ағаштардан аз).



Кездейсоқ ормандар - оларды не үшін пайдалану керек?

- Әрине, деректердегі сызықтық емес қатынастарды өңдеңіз.
- Тез қонады.
- Шамадан тыс орнатуға төзімді (бірақ иммунитет емес).
- Деректеріңізді масштабтаудың қажеті жоқ.
- Олар түсіндірілетін (түрі) болуы мүмкін. See [feature importances](#) which measures the total Gini reduction brought by each feature.
- Олар шынымен жақсы жұмыс істейді!