**№2 Дәріс тақырыбы: Модель, супермодель, шешім ағашы**

**Кіріспе:**

Модельдер өмірдің көптеген салаларында қолданылады. Инженерлер құрылыс жобаларын модельдейді, ғалымдар физикалық процестерді модельдейді, ал деректер ғалымдары мәліметтерге негізделген шешім қабылдау үшін **математикалық модельдер** құрады. Бүгінгі дәрісімізде біз машиналық оқытудағы ең танымал модельдердің бірі – **шешім ағашы (Decision Tree)** туралы сөйлесеміз. Шешім ағашы – қарапайым әрі интуитивті түсінікті модель, сондықтан біз "Модель, жоқ, супермодель емес, ағаштан бастайық" деп айтамыз.

**1. Модель дегеніміз не?**

Алдымен, **модель** терминіне қысқаша тоқталайық. Модель – бұл нақты объектіні, процесті немесе жүйені белгілі бір түрде сипаттайтын математикалық немесе логикалық түсінік. Модельдің негізгі мақсаты – деректерді талдау және болжау жасау.

**Модельдің негізгі сипаттамалары:**

* **Деректер негізінде құрылған:** Модель өзіне берілген деректерге сүйене отырып, белгілі бір заңдылықтар мен байланыстарды анықтайды.
* **Шешім қабылдау мүмкіндігі:** Модельдің көмегімен болашақтағы жағдайларды болжауға немесе шешім қабылдауға болады.
* **Оқытылатын:** Машиналық оқыту модельдері берілген деректер негізінде үйренеді және болашақта ұқсас жағдайлар үшін болжам жасайды.

**2. Шешім ағашы (Decision Tree) дегеніміз не?**

**Шешім ағашы** – қадағаланатын оқытуда (Supervised Learning) қолданылатын алгоритм, ол деректер негізінде шешім қабылдау процесін модельдейді. Бұл алгоритм, оның аты айтып тұрғандай, ағашқа ұқсас құрылым жасайды, мұнда әрбір түйін белгілі бір сұрақты (шартты) білдіреді, ал тармақтар (branches) сол сұраққа берілген жауаптарға сәйкес келетін нәтижелерді көрсетеді.

**2.1 Шешім ағашының құрылымы:**

* **Түбірлік түйін (Root Node):** Шешім ағашының бастапқы нүктесі. Бұл түйінде деректердің барлық жиынтығы талданады.
* **Ішкі түйіндер (Internal Nodes):** Әрбір түйінде сұрақ немесе логикалық шарт тұрады, мысалы: «Бұл жасы 30-дан үлкен ме?».
* **Жапырақтар (Leaf Nodes):** Бұл түйіндерде жауап немесе соңғы шешім беріледі. Мысалы, сыныптау мәселесінде жапырақтар деректердің қай сыныпқа жататынын көрсетеді.

**2.2 Шешім ағашының жұмысы:**

Шешім ағашы шешім қабылдау процесін шарттарға бөлуден бастайды. Әрбір ішкі түйінде бір айнымалының мәні бойынша шешім қабылданады. Мысалы, егер сіздің мақсатыңыз белгілі бір адамдардың табыстарын болжау болса, алғашқы түйінде адамның жасына, содан кейін білім деңгейіне қарай шешім қабылдайтын шарттар болуы мүмкін.

**2.3 Шешім ағашының артықшылықтары:**

* **Түсінікті:** Шешім ағашы қарапайым және интуитивті түрде түсінікті, өйткені ол шешім қабылдау процесін көрнекі түрде көрсетеді.
* **Шешімдерді түсіндіру:** Шешім ағашының әрбір қадамы түсінікті болғандықтан, ол қабылдаған шешімдерді оңай түсіндіруге болады.
* **Құрылымдылығы:** Ағаш құрылымы шешім қабылдау процесін белгілі бір логикаға сай құрады, бұл деректерді автоматты түрде өңдеуге мүмкіндік береді.

**2.4 Шешім ағашының кемшіліктері:**

* **Ағаштың үлкендігі:** Шешім ағашы тым күрделі болса, ол «артық үйрену» (overfitting) жағдайына алып келуі мүмкін.
* **Деректердегі аздап өзгерістер:** Шешім ағашының құрылымы аз ғана өзгерістер кезінде өзгеріп кетуі мүмкін, бұл нәтижелердің тұрақтылығына әсер етуі мүмкін.

**3. Шешім ағашын құру алгоритмі**

**3.1 Шешім ағашын қалай құрады?**

Шешім ағашын құру үшін деректер жиынындағы әртүрлі шарттарды тексеріп, ең тиімді шешімдерді қабылдау қажет. Ең жиі қолданылатын әдістердің бірі – **Энтропия** мен **Ақпараттық пайда** (Information Gain).

* **Энтропия (Entropy):** Деректер жиынындағы белгісіздік деңгейін көрсететін шама. Энтропия неғұрлым аз болса, мәліметтер соғұрлым біржақты болады.
* **Ақпараттық пайда (Information Gain):** Энтропияның төмендеу деңгейін анықтайды. Егер шарт бір айнымалының энтропиясын азайтса, онда бұл айнымалы ағашта жоғарырақ тұрады.

**3.2 Шешім ағашын құру қадамдары:**

1. **Барлық деректерді түбірлік түйінге қою:** Алғашқы шарт түбірлік түйінде орналасады.
2. **Айнымалыны таңдау:** Келесі шарт қандай айнымалыға қатысты болатынын таңдау керек. Бұл ақпараттық пайданы есептеу арқылы жүзеге асады.
3. **Түйіндерді бөлу:** Шартқа сәйкес деректер екі немесе одан да көп топқа бөлінеді.
4. **Қайталау:** Әрбір жаңа топ үшін жаңа түйіндерге шарттар қойылады. Бұл процесс барлық деректер жапырақтарға жеткенше жалғасады.

**3.3 Шешім ағашын оңтайландыру:**

Артық үйрену мәселесін шешу үшін ағашты дұрыс өлшемде ұстау маңызды. Кейбір әдістер ағашты «кесуді» (pruning) қолданады, яғни, тым көп түйіндер немесе тармақтарды жояды, бұл модельдің күрделілігін азайтады және жалпылау мүмкіндігін жақсартады.

**4. Шешім ағашының қолдану мысалдары**

**4.1 Мысал 1: Жазғы демалыс орнының ұсынысын таңдау**

Клиенттің демалысқа қайда баратынын болжау үшін шешім ағашын құруға болады. Айнымалылар: демалыс түрі (теңіз, тау), бюджеті, және демалыс уақыты. Шешім ағашы осы айнымалыларды шарттарға бөліп, клиентке ең жақсы ұсынысты жасайды.

**4.2 Мысал 2: Медициналық диагноз қою**

Науқастың денсаулығына қатысты мәліметтер (мысалы, жас, қан қысымы, ауру белгілері) негізінде шешім ағашын құрып, дәрігер науқастың қандай ауруға шалдыққанын болжай алады.

**5. Шешім ағашының басқа машиналық оқыту әдістерімен байланысы**

Шешім ағашы – өте қуатты модель болғанымен, ол күрделі мәселелерде жалғыз қолданылмайды. Оның бірнеше вариациялары бар:

* **Random Forest (Кездейсоқ орман):** Көптеген шешім ағаштарын біріктіріп, нәтижесін жақсарту әдісі.
* **Gradient Boosting:** Шешім ағаштарын рет-ретімен құрып, әр ағаш алдыңғы ағаштың қателіктерін түзетуге тырысады.

**Қорытынды:**

Шешім ағашы – машиналық оқытудың қарапайым және түсінікті модельдерінің бірі. Ол нақты шешім қабылдау процестерін көрнекі түрде модельдейді және көптеген қолданбалы салаларда кеңінен қолданылады. Күрделі мәліметтерді талдау және болжау үшін ағаштан бастау дұрыс шешім, себебі ол бастапқыда қарапайым болып көрінгенімен, оның негізінде көптеген тиімді модельдер құруға болады.