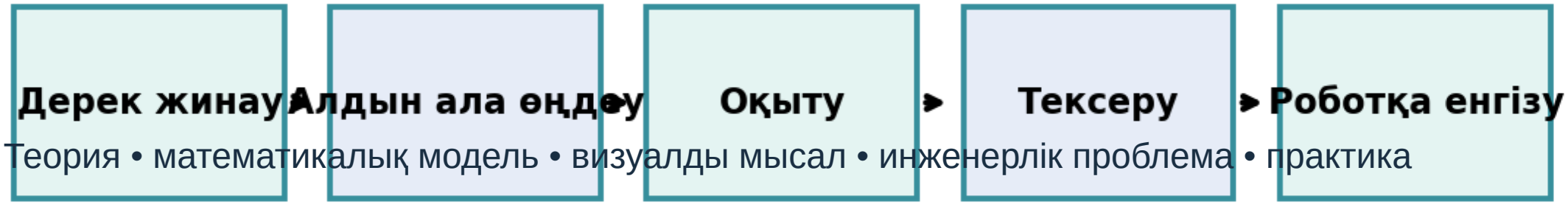
Машиналық оқыту: негізгі алгоритмдер

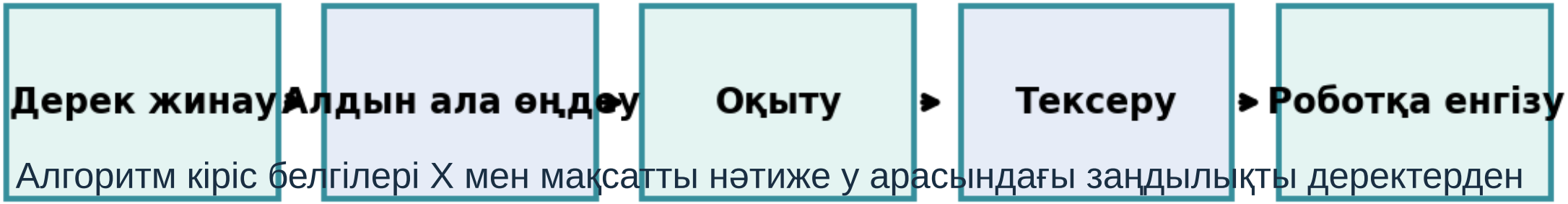
Робототехникадағы деректерге негізделген болжау, тану және топтау



Дәрістің оқу мақсаттары

- Машиналық оқытудың (Machine Learning) негізгі ұғымдарын және түрлерін түсіндіру.
- Бақыланатын және бақыланбайтын оқытудың айырмашылығын ажырату.
- Регрессия, классификация және кластеризация есептерін дұрыс тұжырымдау.
- Алгоритм сапасын MAE, RMSE, Accuracy, Precision, Recall және Silhouette арқылы бағалау.
- Роботтың сенсорлық деректері үшін әдісті таңдап, негізгі тәуекелдерді анықтау.

Машиналық оқыту қалай жұмыс істейді?



Алгоритм кіріс белгілері X мен мақсатты нәтиже y арасындағы заңдылықты деректерден үйренеді. Оқыту жиыны модель параметрлерін анықтайды; валидация гиперпараметрлерді таңдауға көмектеседі; тест жиыны жалпылау сапасын бағалайды.

Оқыту түрлерінің айырмашылығы

Supervised learning

Белгіленген деректер бар: регрессия және классификация. Мысалы, мотор температурасы арқылы ақауды болжау.

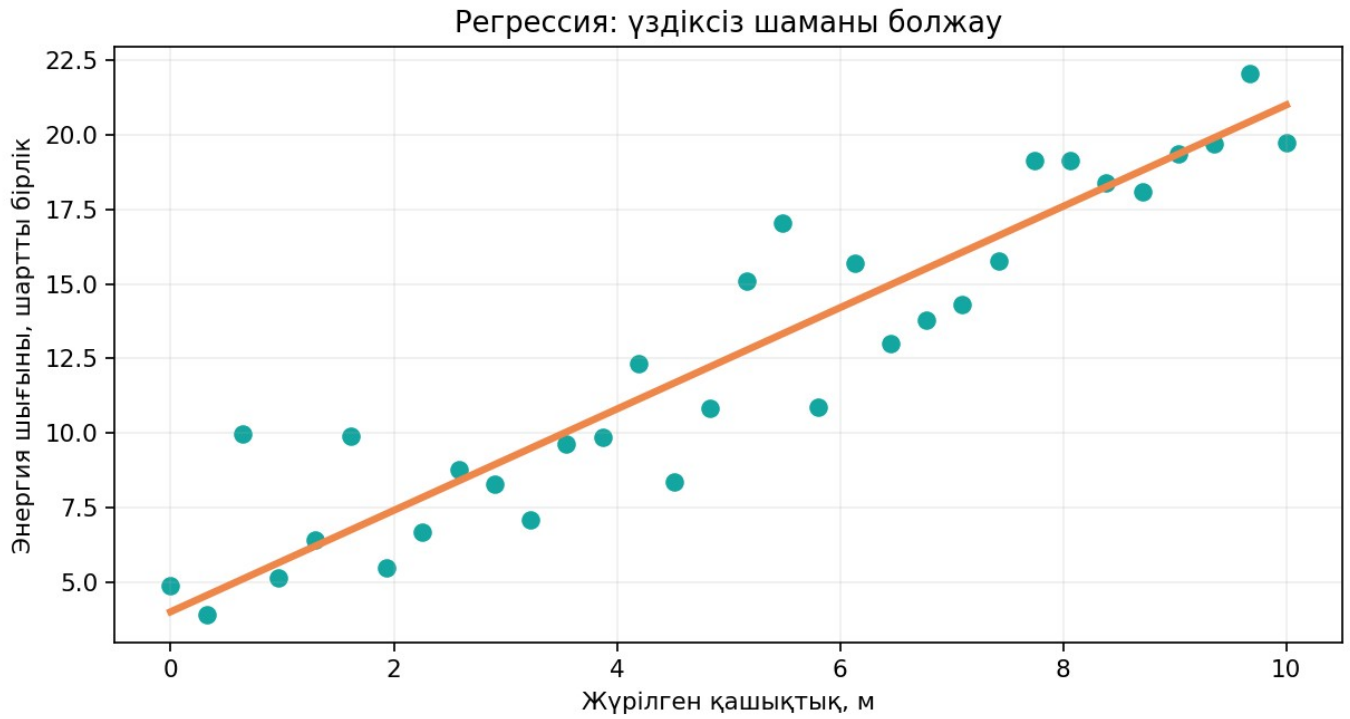
Unsupervised learning

Дайын жауап жоқ: кластеризация және құрылымды іздеу. Мысалы, сенсорлық режимдерді топтау.

Reinforcement learning

Агент әрекет жасап, сыйақы бойынша саясатты үйренеді. Мысалы, робот қозғалысын оңтайландыру.

Регрессия: сандық шаманы болжау



Регрессияның шығысы — үздіксіз сан: қашықтық, температура, қуат, ресурс.

Сызықтық модель: $\hat{y} = w_0 + w_1 x$.

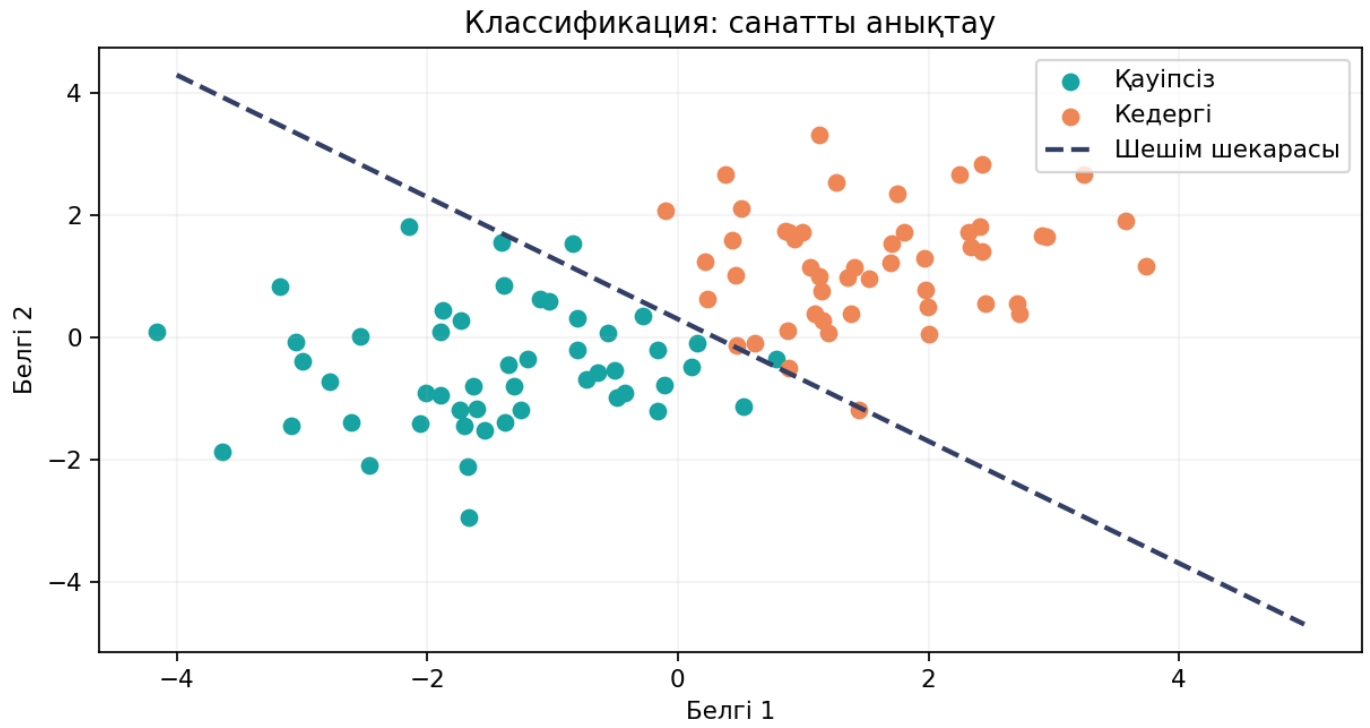
Әдістер: Linear Regression, Ridge, Random Forest Regressor.

Робот мысалы: аккумулятор қалдығы бойынша жұмыс уақытын болжау.

Регрессия: сапа және проблемалар

- MAE = орташа абсолют қате; RMSE = квадраттық қателердің орташа мәнінің түбірі.
- Артық үйрену (overfitting): оқыту дерегінде жақсы, жаңа ортада нашар нәтиже.
- Сенсор шуылы мен өлшем бірліктерінің сәйкес келмеуі болжам дәлдігін төмендетеді.
- Шешімдер: деректерді тазалау, кросс-валидация, тұрақтандыру және тәуелсіз тест жиыны.
- Роботта болжам қатесі қауіпсіздік шегінен асса, резервтік басқару алгоритмі қажет.

Классификация: санатты анықтау



Шығыс — дискретті класс: «кедергі/жол бос», «ақау/қалыпты», «объект түрі».

Алгоритмдер: Logistic Regression, Decision Tree, SVM, k-NN, нейрондық желілер.

Шешім шекарасы белгілер кеңістігін кластарға бөледі. Нақты әлемде кластар жиі қабаттасады.

Классификация сапасын бағалау

Accuracy

Барлық дұрыс болжамдардың үлесі. Кластар теңгерімсіз болса, жаңылыстыруы мүмкін.

Precision

«Қауіпті» деген болжамдардың қаншасы шынымен қауіпті екенін көрсетеді.

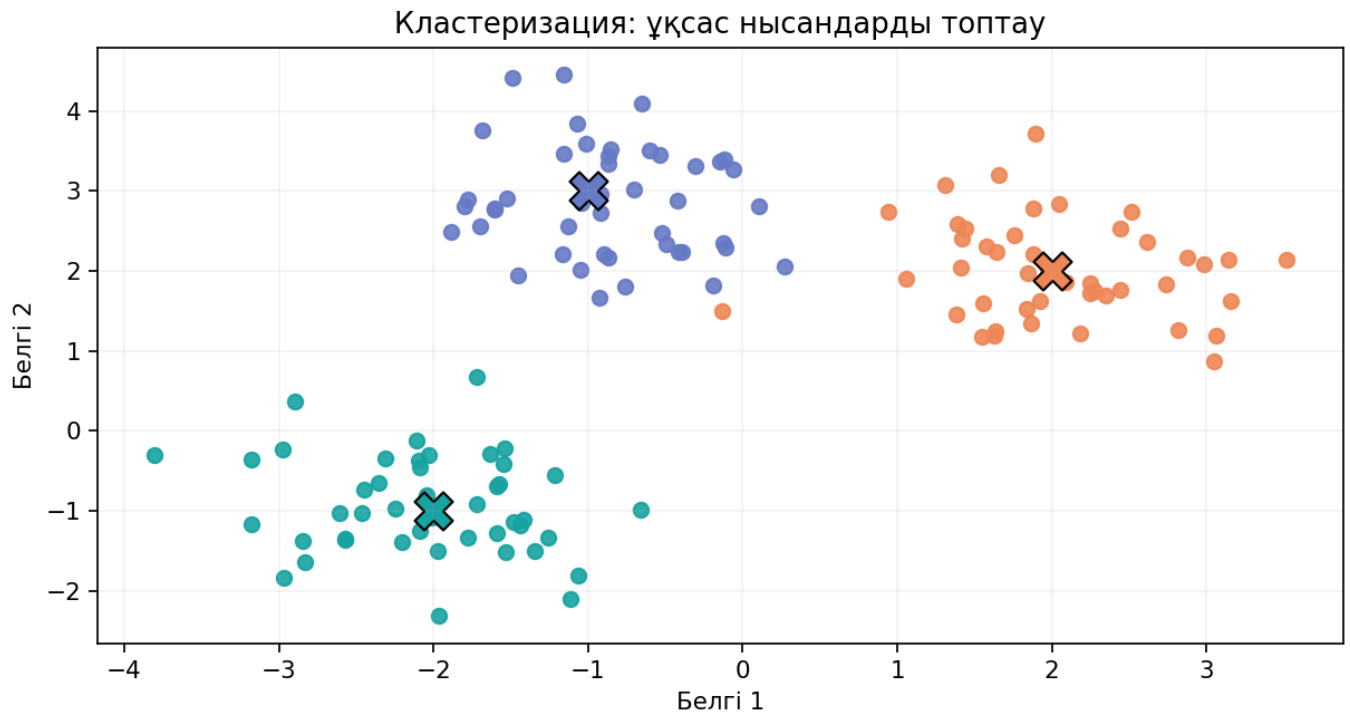
Recall

Нақты қауіпті жағдайлардың қаншасын анықтағанын көрсетеді.

F1-score

Precision мен Recall арасындағы үйлесімді орташа көрсеткіш.

Кластеризация: белгісіз құрылымды табу



Кластеризация белгілері ұқсас нысандарды алдын ала белгісіз топтарға бөледі.

K-means: K орталыққа дейінгі қашықтықты азайтады.

DBSCAN: тығыз топтарды табады, шуды бөліп көрсете алады.

Робот мысалы: лидар нүктелерін ықтимал объектілерге топтау.

Кластеризациядағы инженерлік қиындықтар

- K-means үшін K саны алдын ала беріледі; қате K топтардың құрылымын бұрмалайды.
- Масштабы әртүрлі белгілер қашықтық өлшеміне шамадан тыс әсер етеді.
- Шу, сирек нүктелер және әртүрлі тығыздық кластерлерді ажыратуды қиындатады.
- Шешімдер: StandardScaler, Silhouette талдауы, DBSCAN және сапаны визуалды тексеру.
- Кластерлер міндетті түрде физикалық объектілермен сәйкес келмейді: қосымша валидация қажет.

Үш алгоритмді салыстыру

Критерий	Регрессия	Классификация	Кластеризация
Нәтиже	Сан	Класс	Топ идентификаторы
Белгіленген у	Қажет	Қажет	Қажет емес
Мысал	Қуатты болжау	Ақауды анықтау	Лидар топтары
Метрика	MAE, RMSE	F1, Recall	Silhouette
Тәуекел	Болжау ауытқуы	Жалған шешім	Жалған топ

Робототехникадағы біріктірілген мысал

Міндет: қоймадағы автономды робот кедергілерді анықтап, батарея ресурсын жоспарлауы тиіс.

1. Классификация: камера кадрынан адам немесе жүк арбасын тану.
2. Кластеризация: лидар нүктелерін кеңістіктік объектілерге бөлу.
3. Регрессия: маршрут ұзындығы мен жүктеме арқылы энергия шығынын болжау.

Бірнеше модельдің нәтижесі қауіпсіз қозғалыс жоспарлағышына беріледі.

Деректер сапасы және қауіпсіздік проблемалары

- Dataset shift: жарық, ауа райы немесе жаңа орта өзгерсе, модель сапасы төмендеуі мүмкін.
- Class imbalance: сирек қауіпті оқиғалар аз кездескендіктен, оларды тану қиындайды.
- Data leakage: тест дерегінің оқытуға араласуы шынайы сапаны асыра көрсетеді.
- Нақты уақыт: модельдің кідірісі тежеу мен кедергіден айналып өтуге әсер етеді.
- Шешімдер: мониторинг, қайта оқыту, fail-safe режимі, түсіндірілетін метрикалар және тестілеу.

Практикалық тапсырма және бақылау сұрақтары

- Робот сенсорларынан 1000 жазбалық деректер жиынын жобалаңыз: жылдамдық, ток, қашықтық, температура.
- Қай мақсатқа регрессия, классификация және кластеризация қолданатыныңызды түсіндіріңіз.
- Деректерді train/validation/test бөліктеріне бөліп, әр модельге сәйкес метрика таңдаңыз.
- «Ақау» класы тек 2% болса, неге Accuracy жеткіліксіз екенін дәлелдеңіз.
- Нәтижені жаңа ортада тексеру және қауіпсіздікке әсерін бағалау жоспарын жасаңыз.

Қорытынды және әдебиеттер

Регрессия сандық шамаларды болжайды; классификация нысандарды санаттарға бөледі; кластеризация белгісіз құрылымды ашады. Робототехникада олардың сапасы тек алгоритмге емес, деректерге, сенсорларға, есептеу кідірісіне және қауіпсіздік талаптарына тәуелді.

Әдебиеттер: Bishop, Pattern Recognition and Machine Learning; Géron, Hands-On Machine Learning; scikit-learn User Guide (supervised және unsupervised learning).