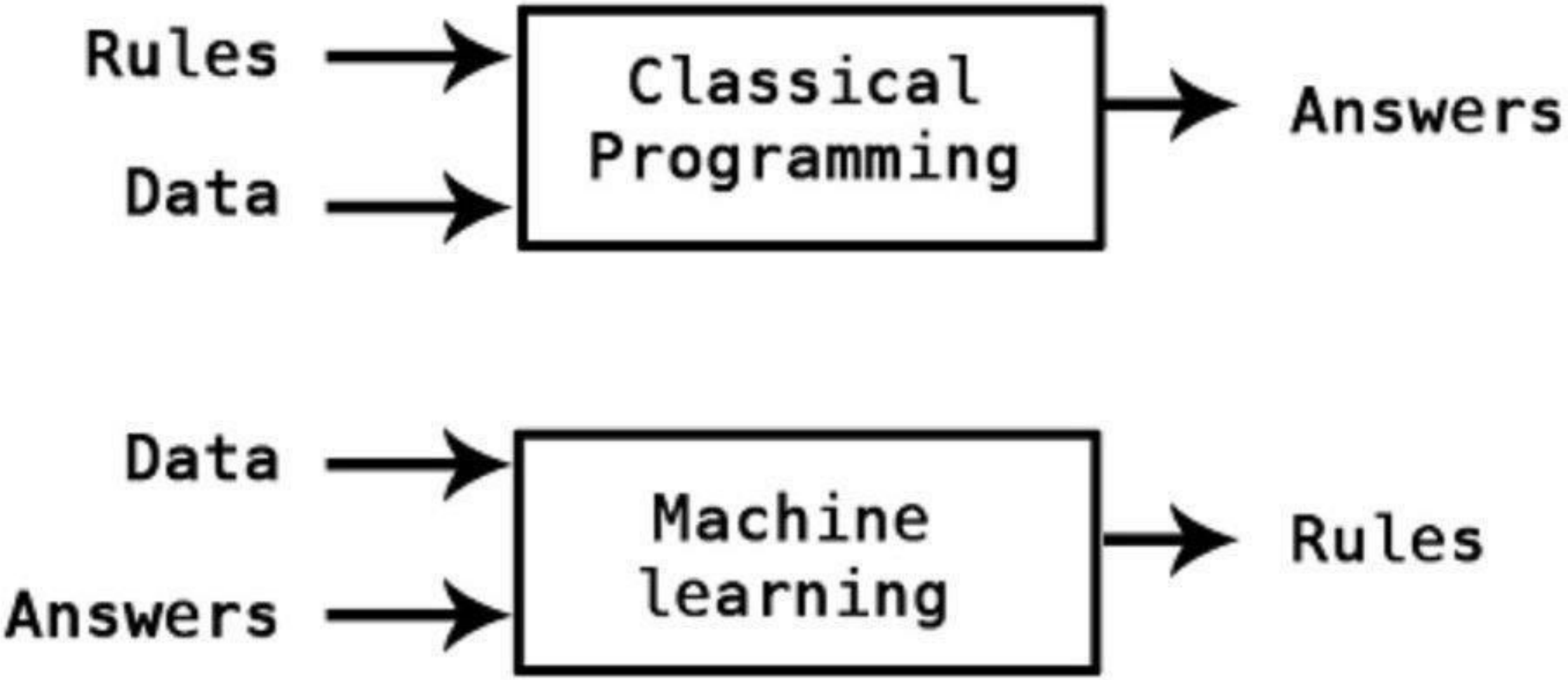
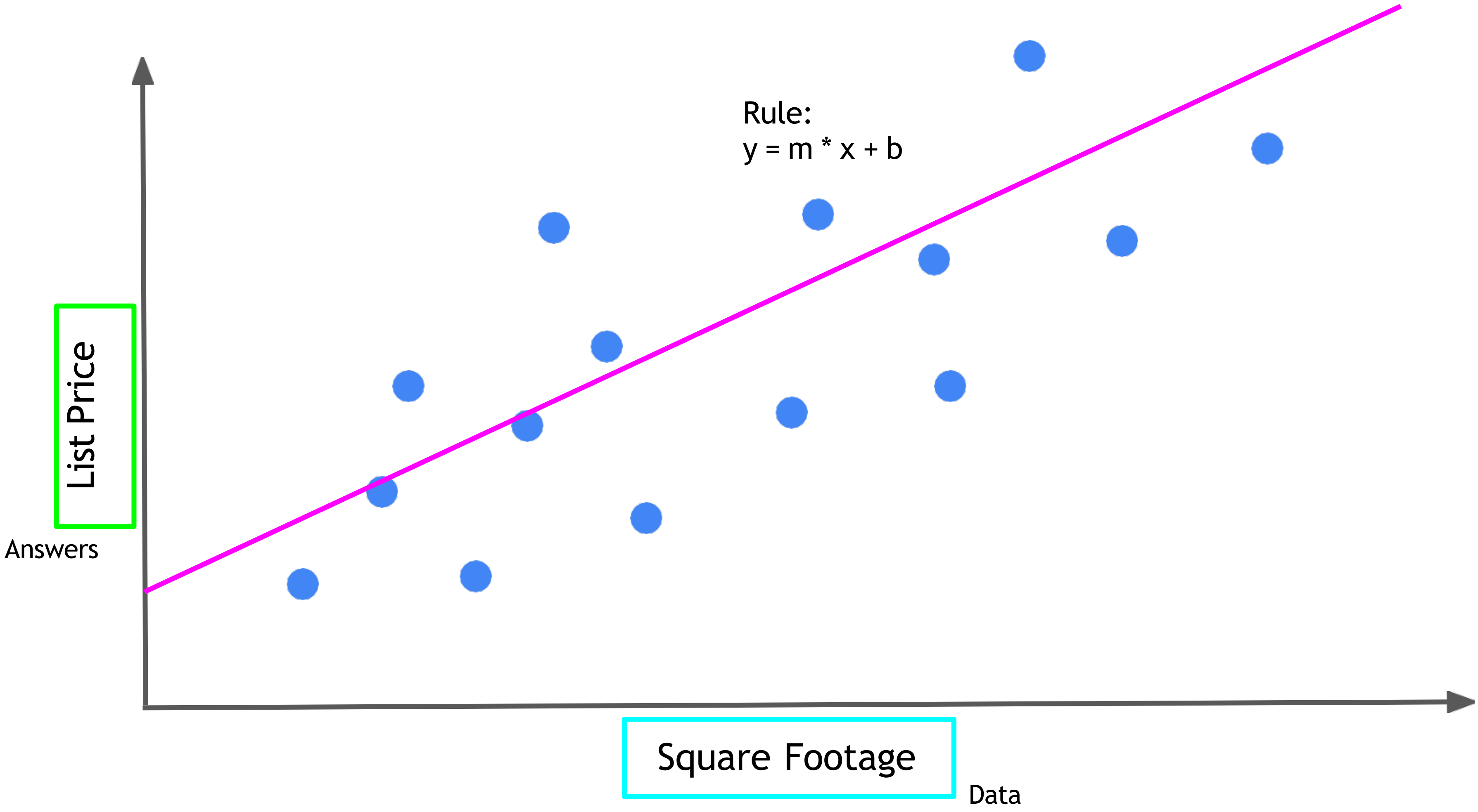
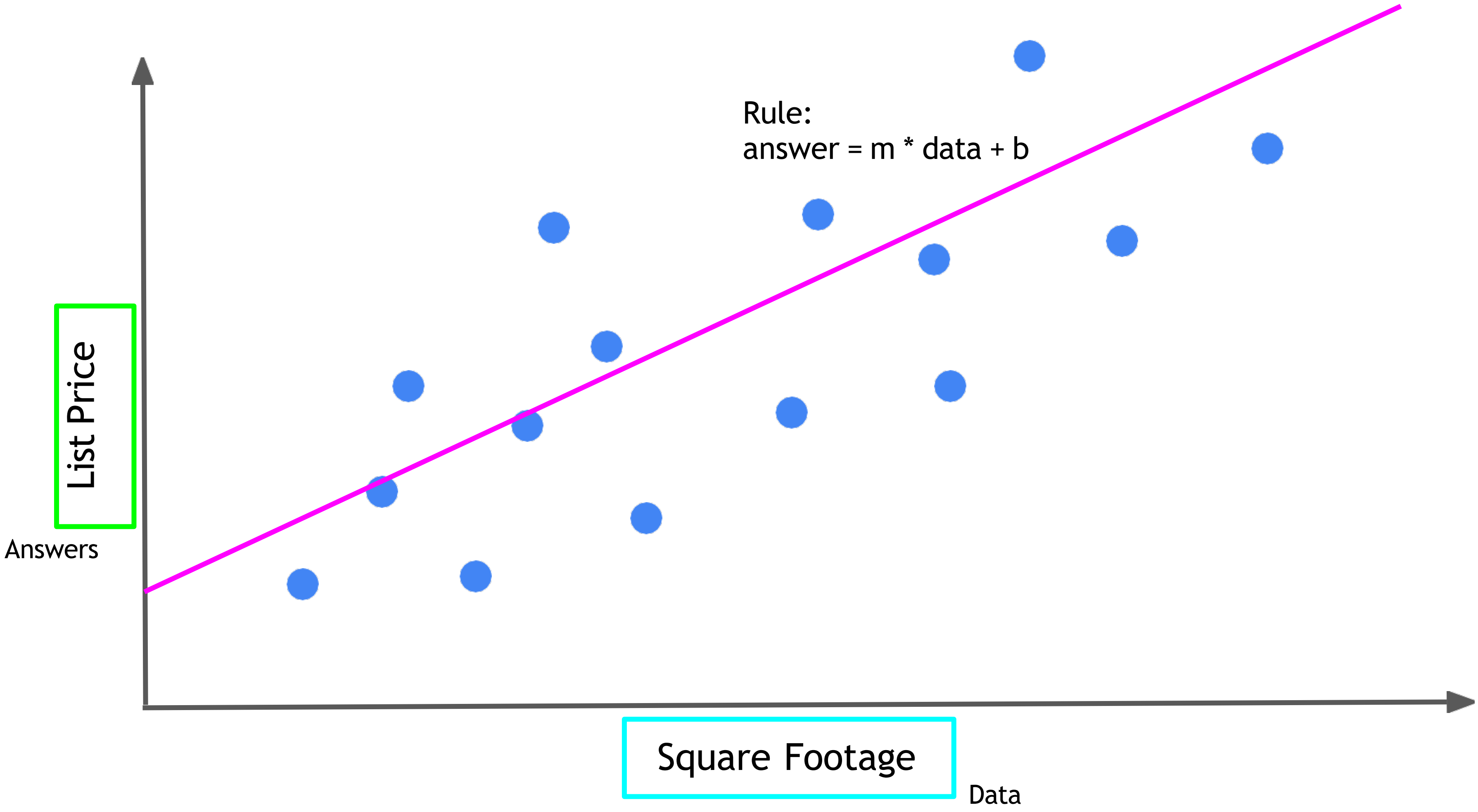
Регрессияға арналған сызықтық модельдер

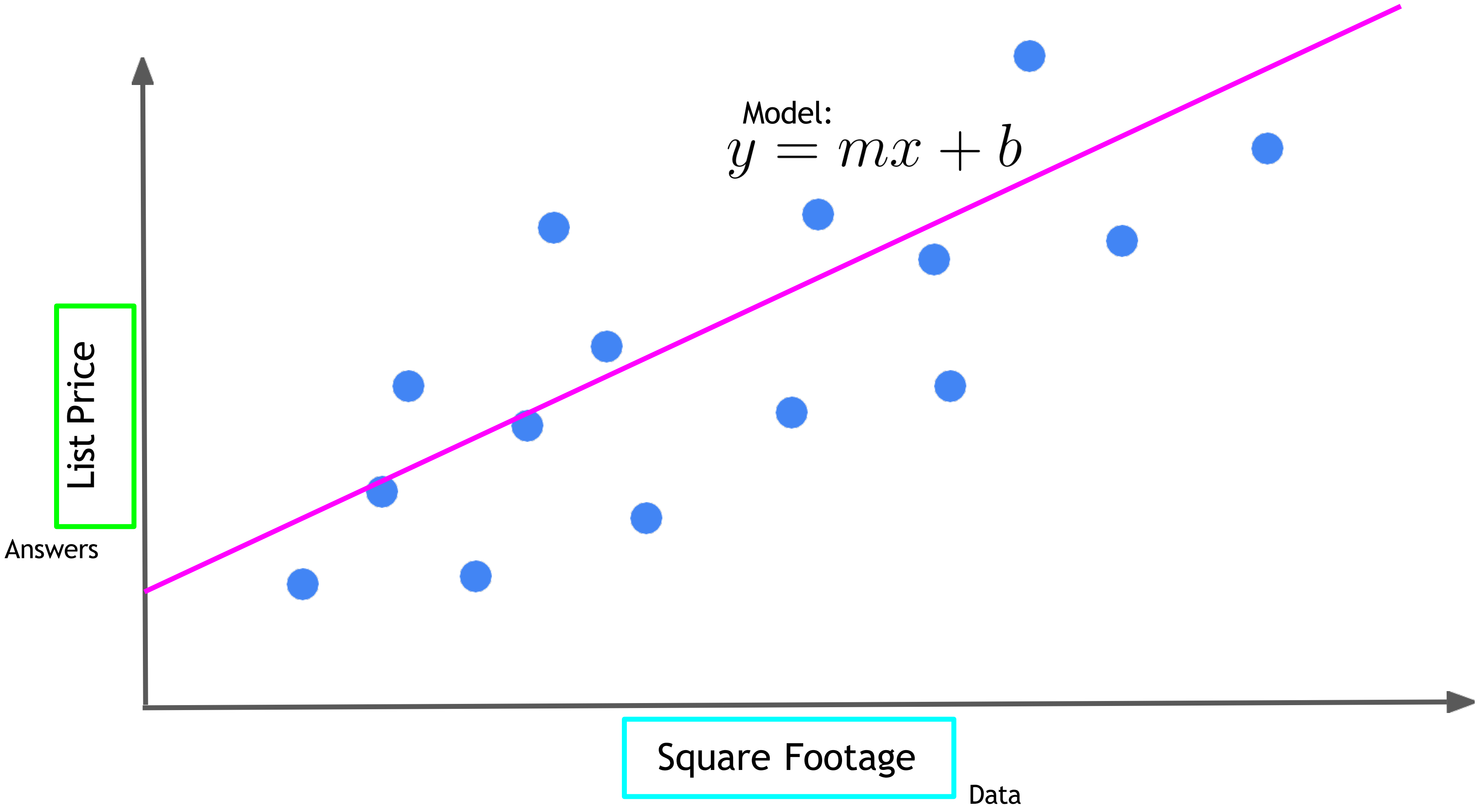
Регрессияға арналған сызықтық модельдер

- **Регрессия:** үздіксіз мәндерді болжау кезінде.
- **Сызықтық модель:** модель параметрлерінде/коэффиценттерінде сызықтық болатын модель.


Rules
Data
Classical Programming
Answers
Data
Answers
Machine learning
Rules


Rule:
y = m * x + b
List Price
Answers
Square Footage
Data


Rule:
answer = m * data + b
List Price
Answers
Square Footage
Data


Model:
$y = mx + b$
List Price
Answers
Square Footage
Data

Жалпылап алайық

- Бізде і арқылы индекстелген n үлгі немесе деректер нүктесі бар деп елестетіңіз.
- Әрбір үлгі үшін бізде j арқылы индекстелген p мүмкіндіктері (үлгі туралы белгілі «ақпарат бөлігі») бар.
- Сызықтық модель келесідей болады:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + + \beta_p X_{ip}$$

β_j j-модель параметрін/коэффицентін білдіреді.

Жалпылап алайық

- Бізде i арқылы индекстелген n үлгі немесе деректер нүктесі бар деп елестетіңіз.
- Әрбір үлгі үшін бізде j арқылы индекстелген p мүмкіндіктері (үлгі туралы белгілі «ақпарат бөлігі») бар.
- Сызықтық модель келесідей болады:

Белгілі деректер

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}$$

Белгілі жауап

β_j j-модель параметрін/коэффицентін білдіреді.

«Ережелер»
бұл біздің үйренгіміз келетін нәрсе!

Жалпылап алайық

Жалғау мерзімі (y-кесінді)

- Бізде i арқылы индекстелген n үлгі немесе деректер нүктесі бар деп елестетіңіз.
- Әрбір үлгі үшін бізде j арқылы индекстелген p мүмкіндіктері (үлгі туралы белгілі «ақпарат бөлігі») бар.

- Сызықтық модель келесідей болады:

Белгілі деректер

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}$$

Белгілі жауап

β_j j-модель параметрін/коэффицентін білдіреді.

«Ережелер»
бұл біздің үйренгіміз келетін
нәрсе!

Жалпылау - бір үлгі і

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}$$

Жалпылау - бір үлгі і

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}$$

$$y_i = \sum_{j=0}^{p} \beta_j X_{ij}$$

Жалпылау - n үлгіні қарастырайық

$$y_i = \sum_{j=0}^{p} \beta_j X_{ij}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} X_{10} & X_{11} & X_{12} & \dots & X_{1p} \\ X_{20} & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots & \\ X_{n0} & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_p \end{bmatrix}$$

$$\vec{y} = \mathbf{X}\vec{\beta}$$

Жалпылау - n үлгіні қарастырайық $y_i = \sum_{j=0}^{p} \beta_j X_{ij}$

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} X_{10} & X_{11} & X_{12} & \dots & X_{1p} \\ X_{20} & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots & \\ X_{n0} & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_p \end{bmatrix}$$

Known Data (features)

$$\vec{y} = \mathbf{X}\vec{\beta}$$

Ережелер
бұл біздің үйренгіміз келетін нәрсе!

Белгілі жауаптар

ML рецепті

1. Кейбір үлгіні ойлап көріңіз
2. Деректерді үлгіге енгізіп, болжам жасаңыз.
3. Болжамдар мен шынайы мәндер арасындағы шығынды есептеңіз.
4. Ең аз шығын келтіретін үлгі параметрлерін анықтаңыз.

ML рецепті

1. Кейбір үлгіні ойлап көріңіз
2. Деректерді үлгіге енгізіп, болжам жасаңыз.
3. Болжамдар мен шынайы мәндер арасындағы шығынды есептеңіз.
4. Ең аз шығын келтіретін үлгі параметрлерін анықтаңыз.

Модельдік болжамдар жиі белгіленеді $\hat{y}_i$

Негізгі шындық/жауаптар жиі белгіленеді y_i (no hat).

ML рецепті

1. Кейбір үлгіні ойлап көріңіз
2. Деректерді үлгіге енгізіп, болжам жасаңыз.

$$\hat{\vec{y}} = \mathbf{X}\vec{\beta}$$

ML рецепті

1. Кейбір үлгіні ойлап көріңіз
2. Деректерді үлгіге енгізіп, болжам жасаңыз.
3. Болжамдар мен шынайы мәндер арасындағы шығынды есептеңіз.

Орташа квадраттық қате (MSE):

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(\underbrace{y_i}_{\text{шынайы мән}} - \underbrace{\hat{y}_i}_{\text{болжау}})^2$$

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \underbrace{\vec{X}_i}_{\text{деректер}} \cdot \vec{\beta}\right)^2$$

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=0}^{p} X_{ij}\underbrace{\beta_j}_{\text{Unknown!}}\right)^2$$

ML рецепті

1. Кейбір үлгіні ойлап көріңіз
2. Деректерді үлгіге енгізіп, болжам жасаңыз.
3. Болжамдар мен шынайы мәндер арасындағы шығынды есептеңіз.
4. Ең аз шығын келтіретін үлгі параметрлерін анықтаңыз.
    1. Үлгі параметрлеріне қатысты жоғалту функциясының туындысын алыңыз.
    2. Оны нөлге теңестіріңіз.
    3. Модель параметрлерін шешіңіз.

$$\frac{\partial \mathcal{L}}{\partial \vec{\beta}} = 0$$

$$\vec{\beta} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \vec{y}$$

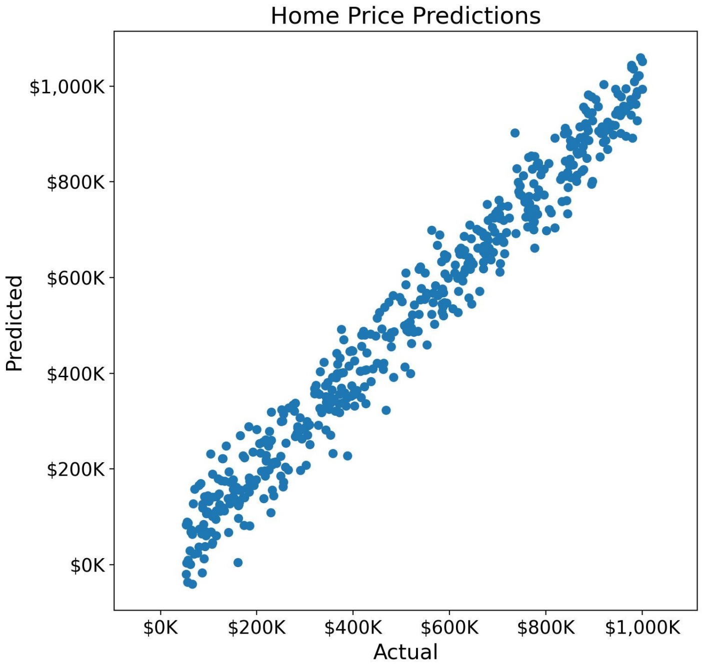
ML рецепті - Мета түсініктемелері

1. Кейбір үлгіні ойлап көріңіз
   a. Біз сызықтық модельді таңдадық. Басқалар да көп!
2. Деректерді үлгіге енгізіп, болжам жасаңыз.
   a. Біз қандай мүмкіндіктерді пайдалану керектігін таңдадық.
3. Болжамдар мен шынайы мәндер арасындағы шығынды есептеңіз.
   a. Біз жоғалту ретінде орташа квадрат қатесін таңдадық. Басқалар да көп!
4. Ең аз шығын келтіретін модель параметрлерін анықтаңыз.
   a. Біз параметрлерді аналитикалық түрде анықтадық. Оларды анықтаудың басқа да көптеген жолдары бар!
   b. Сондай-ақ, кейде параметрлерді аналитикалық түрде анықтай алмайсыз.

1-қадамды «үлгі» деп белгілесек те, бұл қадамдардың барлығы болжау үшін пайдаланылатын түпкілікті үлгіге әсер ететін таңдауларды қамтиды.

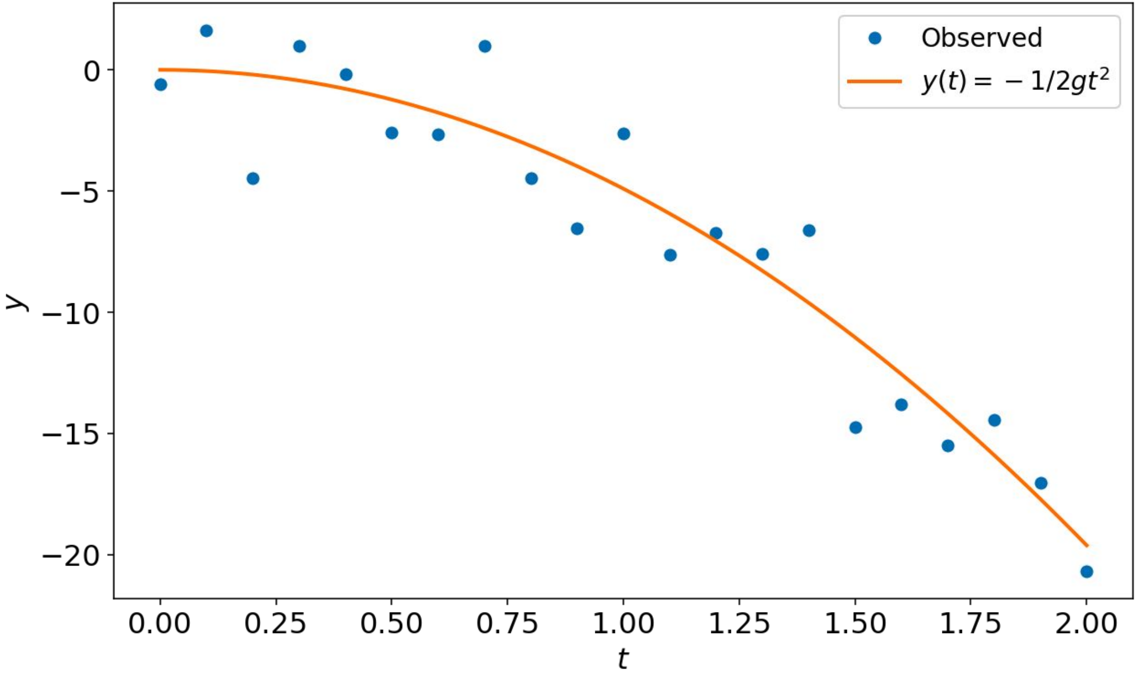
Болжамға(prediction) vs қорытынды(inference)

- Осы уақытқа дейін біз нәрселерді болжайтын модель құра білуге назар аудардық.



Болжамға vs қорытынды

- Қорытынды: жүйеңіз туралы білу үшін үлгіңізді тексеріңіз.



Болжамға vs қорытынды

- Қорытынды: жүйеңіз туралы білу үшін үлгіңізді тексеріңіз.
- Мысалдар:
- Дәл болжау үшін қандай мүмкіндік маңызды?
- Егер Z мүмкіндігін 10%-ға арттырсам, бұл болжамды қалай өзгертеді?
- Менің болжамымдағы белгісіздік неде?
- Үлгі параметріндегі белгісіздік дегеніміз не?
- Бұл көбінесе қиынырақ және статистикалық кепілдіктерді талап етеді.
- Бұл сызықтық модельдер үшін жақсы зерттелген, бірақ күрделі модельдер үшін олай емес.
- Модельдің күрделілігі мен түсіндірмелілігінің арасында өзіндік айырбас бар.
- Модельдің күрделілігі үлгі өнімділігімен корреляциялануы мүмкін, сондықтан өнімділік пен интерпретация арасында айырбас болуы мүмкін

Сызықтық модельдер үшін сызықты емес мүмкіндіктер

- Біз X мүмкіндіктеріне қалағанымызды жасай аламыз.

$$y_i = \sum_{j=0}^{p} \beta_j X_{ij}$$

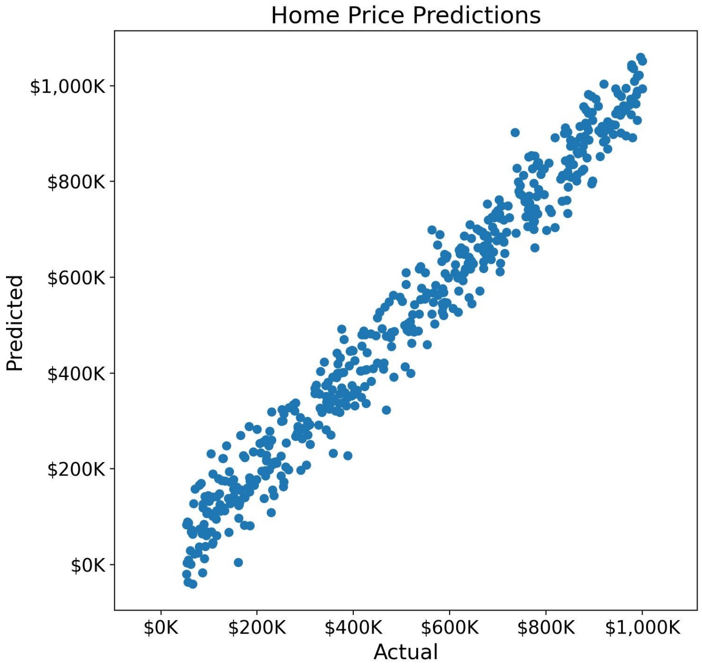
- Біз бір функцияны квадраттай аламыз, біз бір-бірімізбен бірнеше функцияларды жасай аламыз, синусты қолдана аламыз және т.б.

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i1} X_{i2} + \beta_4 sin\left(X_{i3}\right)$$

- Бұл мүмкіндіктер сызықты емес болғанымен, модель параметрлері β бойынша сызықты болады.

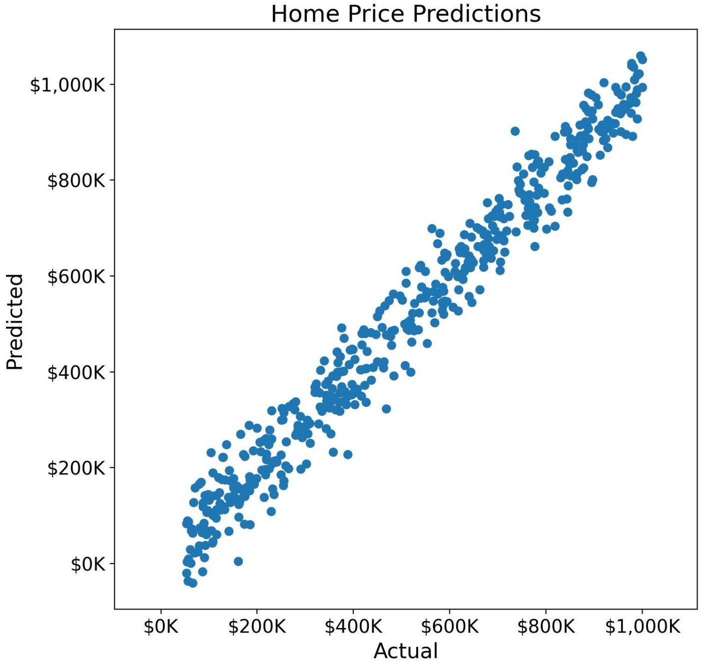
Регрессия үлгілерін бағалау

Біздің болжамымыздың қаншалықты дұрыс екенін қайдан білеміз?



Біздің болжамымыздың қаншалықты дұрыс екенін қайдан білеміз?

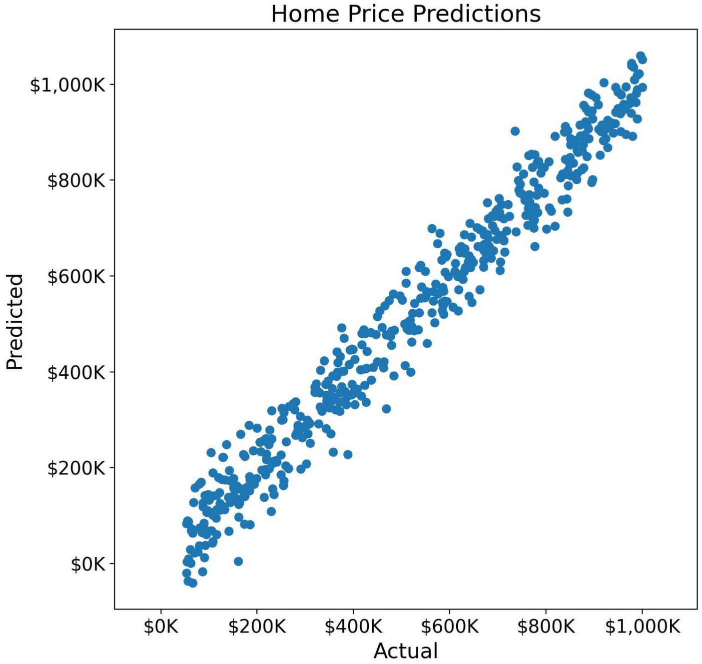
- Модельдің өнімділігі модель қолданылатын контекстке негізделген.
- Біз көбінесе «өнімділікті» бір санға дейін төмендетуге тырысамыз.
- Модель сапасының көптеген статистикалық өлшемдері бар, олар тәжірибеде пайдалы болуы мүмкін немесе болмауы мүмкін.



Біздің болжамымыздың қаншалықты дұрыс екенін қайдан білеміз?

- Егер біз орташа квадраттық қатені (MSE) азайтатын болсақ, онда біз MSE өнімділік көрсеткіші ретінде пайдалана аламыз.
- Бірақ MSE қымбатырақ үйлерді дәлірек болжауға бейім болады.
- Сондай-ақ, шектен тыс мәндер бұл көрсеткішке үлкен әсер етеді.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



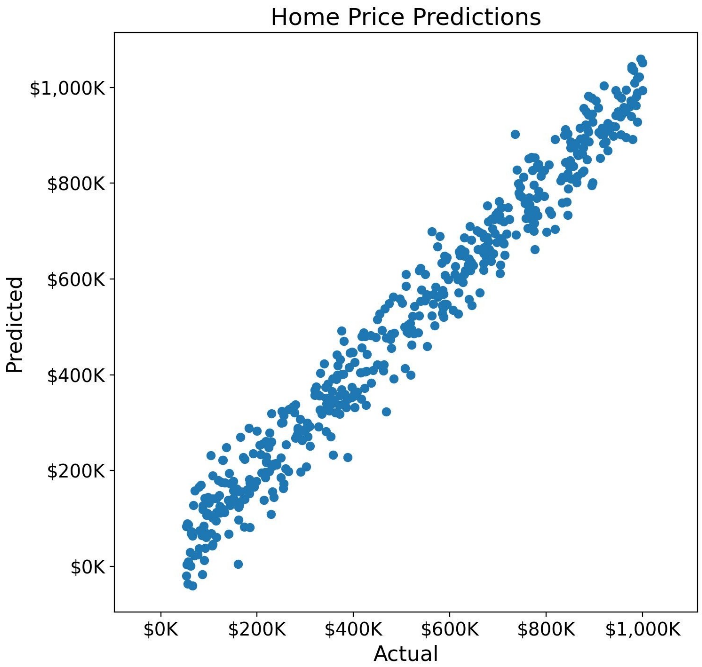
Регрессиялық көрсеткіштер

- Орташа квадрат қатесі - шектен тыс мәндерді азайтады, жақсы статистикалық сипаттарды, түсіндіру қиын.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Орташа абсолютті қате - түсіндіру оңай, барлық шығындар бірдей бағаланады, шектен тыс көрсеткіштерге сенімді емес.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$



Регрессиялық көрсеткіштер

- $\mathcal{R}^2$ / детерминация коэффициенті - болжануға болатын дисперсия үлесі ретінде статистикалық түрде түсіндіріледі

$$\mathcal{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

- Орташа абсолютті пайыздық қате - түсіндіріледі, қате масштабқа тәуелсіз.

$$MAPE = \sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$