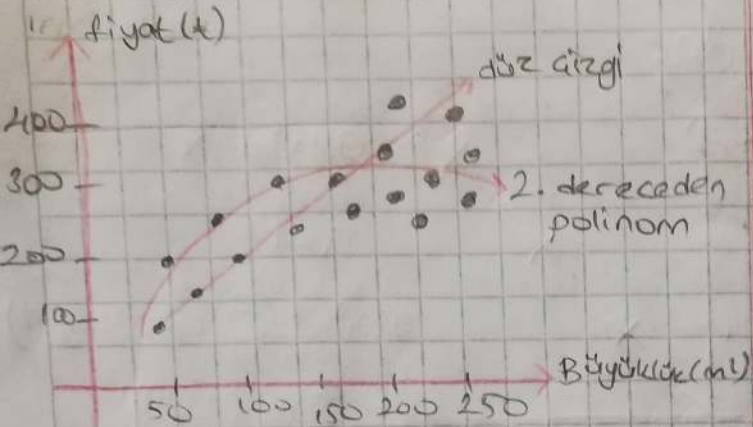
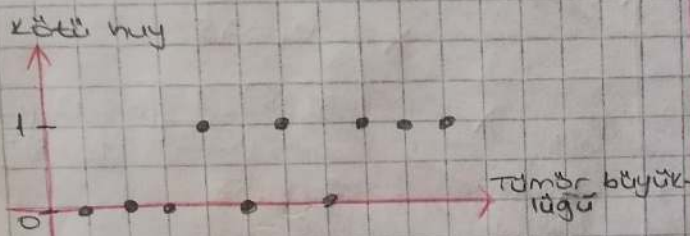


Supervised Learning

* Bilgisayara kendimiz bir şeyler öğrettiğimizde denetimli, bilgisayarın kendi kendine bir şeyler öğrenmesini beklediğimizde denetimsiz öğrenme olur.

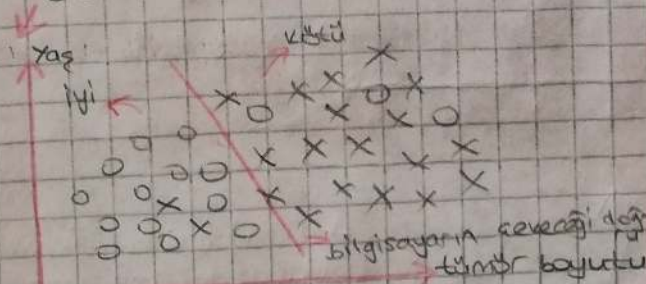


Regresyon problemi: Sürekli değer olan çıktıyı tahmin etmeye çalışır.
 Bizim problem türü



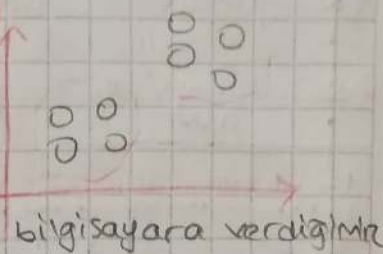
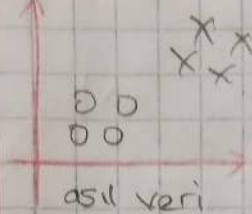
Classification problemi: Kesikli (discrete) çıktıların olasılığını bulma (sınıflandırma) problemidir. Sadece 2 değeri olması gerekmez. Belirli sayıda ve ayrık olması yeterlidir.

0 iyi X kötü



* Bilgisayara ne kadar çok veri girilirse o kadar doğru bir algoritma tahmini almış olur.

Unsupervised Learning



* Bilgisayar elinde sonuçlar ve karşılıkları olmadığı için bu datalar arasında ortak yön bulmaya çalışır ve buna göre kümeleme yapar. Bu algoritmaya (cluster) kümeleme alg. denir.

* Google news de bu algoritmayı kullanır.

Linear Regression:

* Algoritmanın görevi data set'lere en uygun/optimize olan modeli data set'e oturtmaktır.

- training set = data set
- m = number of training examples
- X = input variable / features
- y = output " / target variable
- (x, y) = one training example
- (x_i, y_i) = ith training example

training set

Learning Alg.

size of house (x) → estimated price (y)

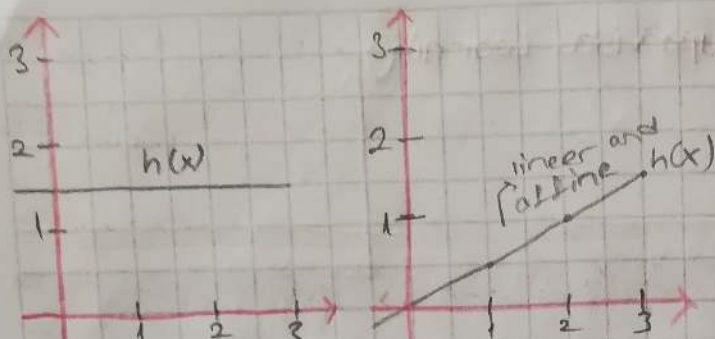


* h maps from x's to y's.

$$h(x) = \theta_0 + \theta_1 \cdot x \quad \text{Linear Regression}$$

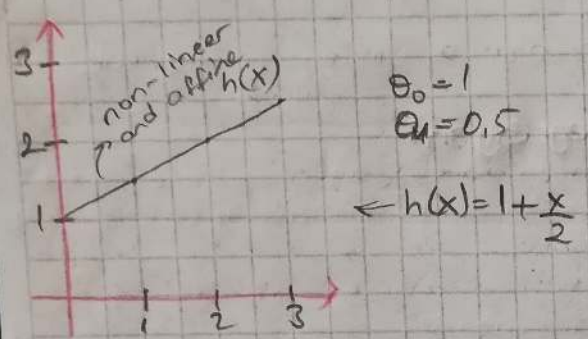
tahmin hipotezi
 $h_\theta(x) = \theta_0 + \theta_1 \cdot x \rightarrow \theta_i$ parameters

* öyle θ_0 ve θ_1 'ler seçeceğiz ki $h_\theta(x)$ y'ye olabildiğince yakın olsun.



$\theta_0 = 1.5$
 $\theta_1 = 0$

$\theta_0 = 0$
 $\theta_1 = 0.5$



* Amacımız θ_0 ve θ_1 'i minimize etmek. aradaki fark negatif olursa denklemi etkilemesin diye karesini alıyoruz

$$\frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

tahmini değer gerçek değer

tüm elemanların hata değeri toplamı
hata değerlerinin ort. bulmak için
yansıyyla işlem yapmak daha kolay olacağı için

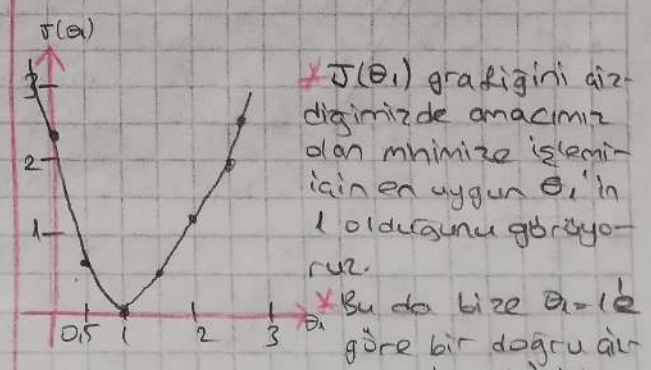
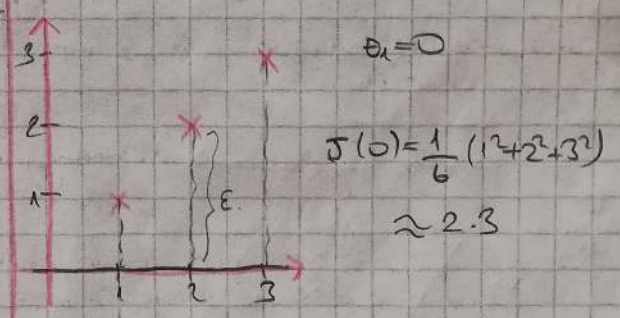
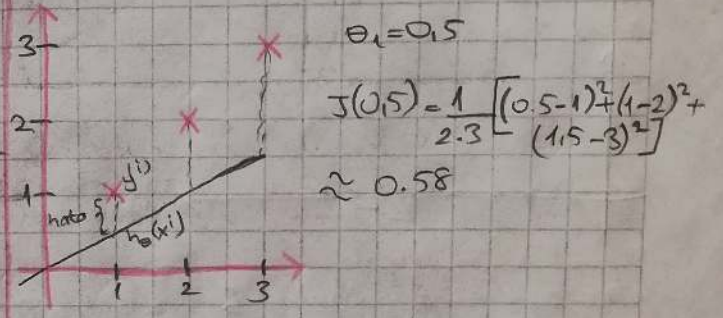
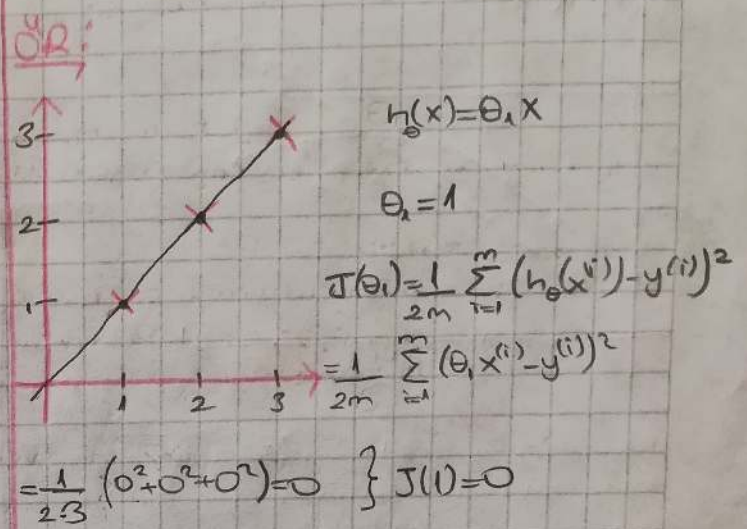
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

minimize $J(\theta_0, \theta_1)$ cost function
squared error func.

- Hypothesis: $h_{\theta}(x) = \theta_0 + \theta_1 x$
- Parameters: θ_0, θ_1
- Cost func.: $J(\theta_0, \theta_1)$
- Goal: minimize $J(\theta_0, \theta_1)$

* Hipotez fonksiyonunu basitleştirmek için θ_0 kadar öteleyeceğiz. Yani fonk. orijinden geçecek ve $\theta_0 = 0$ olacak. Böylece J fonk. da tek en uygun grafik olduğunu gösterir. değişkenli olmuş olacak.

Simplified $h_{\theta}(x) = \theta_1 x$
 $J(\theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$
 goal: minimize $J(\theta_1)$



* $J(\theta_1)$ grafiğini çizdiğimizde amacımız olan minimize işlemin için en uygun θ_1 'in 1 olduğunu görüyoruz.
 * Bu da bize $\theta_1 = 1$ göre bir doğru alır nemizin $h_{\theta}(x)$ için

* Önceki kısımda daha iyi anlamak için θ_0 'ı sıfırlamıştık ve iki boyutlu convex bir $J(\theta_0, \theta_1)$ fonk. elde etmiştik.

* Şimdi θ_0 'ı da dahil ettiğimizde iki boyutlu convex değil 3 boyutlu convex bir $J(\theta_0, \theta_1)$ fonk. elde edeceğiz.

Gradient Descent

* Bu yöntem de bizim J fonk. azaltmak için kullanacağımız bir algoritma. J 'nin 2 parametresi olmasına gerek yok. n tane parametre olabilir.

* θ_0 ve θ_1 değerlerini bir değerden başlatıp adım adım azaltarak bir minimum değer bulmaya çalışacağız.

$$\theta_{j+1} = \theta_j - \alpha \cdot \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_j} \quad (\text{for } j=0 \text{ and } j=1)$$

learning rate

azaltacağımız miktar, ne kadar küçükse o kadar küçük adımlarla ilerleriz

NOT: α := atama işareti olarak geri döner.

* Bu işlem θ_0 ve θ_1 için aynı anda ayrı ayrı yapılmalı.

$$\theta_0 := \theta_0 - \alpha \cdot \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} \quad \theta_0 \text{ 'a göre kısmi türev}$$

$$\theta_1 := \theta_1 - \alpha \cdot \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_1} \quad \theta_1 \text{ 'e göre kısmi türev}$$

* Başladığımız θ noktası zaten optimum nokta ise türev 0 çıkacağından θ değişmez.

* İlk başladığımız noktadan ikinci noktaya geldiğimizde eğim öncekinden daha az (mutlak değerce) olacağı için adım ilerledikçe θ_0 ve θ_1 in yeni değeri arasındaki fark da azalır. Bu yüzden learning rate'i değiştirmeye gerek yoktur.

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \cdot \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^i) - y^i)^2$$

$$= \frac{\partial}{\partial \theta_j} \cdot \frac{1}{2m} \sum_{i=1}^m (\theta_0 + \theta_1 x^i - y^i)^2$$

devamı yan tarafta

$j=0$ iken θ_0 'a göre kısmi türev =

$$\frac{1}{m} \sum_{i=1}^m (h_\theta(x^i) - y^i) = A$$

gösterimi kolaylaştırmak için böyle yaptım

$j=1$ iken θ_1 'e göre kısmi türev =

$$\frac{1}{m} \sum_{i=1}^m (h_\theta(x^i) - y^i) \cdot x^i = B$$

$$\left. \begin{aligned} \theta_0 &:= \theta_0 - \alpha \cdot A \\ \theta_1 &:= \theta_1 - \alpha \cdot B \end{aligned} \right\} \text{eş zamanlı güncellenir.}$$

* Convex olan şekillerde ilk adım yönümüz ulaşacağımız noktayı etkilemez çünkü lokal minimum yoktur, ama şekil convex değilse bu durum değişir.

Batch gradient descent: Gradient descent her iterasyonda (adım) bütün m'leri tekrar kontrol eder. Bu da bize maliyet artışı olarak geri döner.

* Tek hipotez için:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \cdot \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \quad y_i = 1 \cdot \theta_0 + \theta_1 \cdot x_i$$

* Birden fazla hipotez için:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \cdot \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 \\ \theta_0 & \theta_1 & \theta_2 \\ \theta_0 & \theta_1 & \theta_2 \\ \theta_0 & \theta_1 & \theta_2 \end{bmatrix} = \begin{bmatrix} y_1 & y_1 & y_1 \\ y_2 & y_2 & y_2 \\ y_3 & y_3 & y_3 \\ y_4 & y_4 & y_4 \end{bmatrix}$$

Multiple Features

size (m ²)	number of bedrooms	floors	age of home	price
x_1	x_2	x_3	x_4	y

n: number of features

$x^{(i)}$: input of i^{th} training example

$x_j^{(i)}$: value of feature j in i^{th} training example

x_0 var ise 1 olarak kabul edilir.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

- hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$

- parameters: $\theta_0, \theta_1, \theta_2, \dots, \theta_n$

- cost func: J

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$$

- for gradient descent:

$$\theta_j := \theta_j - \alpha \cdot \frac{\partial J(\theta_0, \dots, \theta_n)}{\partial \theta_j}$$

* her θ_j için $(\theta_1, \dots, \theta_n)$ arasında tek tek aynı anda atama yapılır.

$$\theta_j := \theta_j - \alpha \cdot \frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$$

(j 0'dan n'e kadar her tek aynı anda güncellenir. j=0'ken $x_0=1$ olur.)

Polynomial Regression

$$h_\theta(x) = \theta_0 + \theta_1 \cdot \text{frontage}_{x_1} + \theta_2 \cdot \text{depth}_{x_2}$$

iyileştirme

$$x = \text{frontage} \cdot \text{depth}$$

$$h_\theta(x) = \theta_0 + \theta_1 \cdot x$$

* integraldeki değişken değiştirmeye benzer bir değişim yapılır.

Normal Equation:

* Adımlarla uğraşmadan optimal θ değerini bulmayı sağlar.

* Kısmi türev uğraşmadan yapılır.

$$X = \begin{bmatrix} 1 & x_1^1 & x_2^1 & x_3^1 \\ 1 & x_1^2 & x_2^2 & x_3^2 \\ 1 & x_1^3 & x_2^3 & x_3^3 \\ 1 & x_1^4 & x_2^4 & x_3^4 \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

$m \times (n+1)$

$$\theta = (X^T X)^{-1} X^T y$$

m dimensional vector

* Gradient Descent dezavantajları

- α seçmek gerekir. (Extra maliyet)
- Needs many iterations (")

* Gradient avantaj

- Works well even when n is large

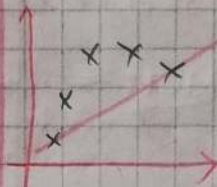
* Normal equation avantaj

- α 'ya gerek yok
- İterasyon yok, tek seferde bulunuyor.

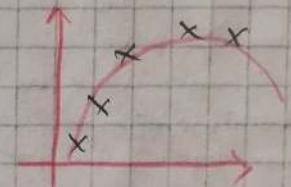
* Normal equation dezavantaj

- Need to compute $(X^T X)^{-1} \rightarrow N^3$ 'lük bir yük demek
- Slow if n is very large.

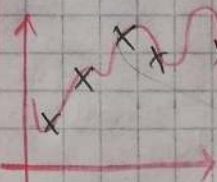
Overfitting



$\theta_0 + \theta_1 x$
- underfitting
- High bias



$\theta_0 + \theta_1 x + \theta_2 x^2$
- just right



$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$
- overfitting
- high variance

* verilen data'lar için hatanın 0'a yakın oldu-

ğu ama yeni örneklerin mevcut fonk. a uymaması

Addressing overfitting

* Özellik sayısı az olduğunda grafik çizilerek overfitting olup olmadığı kontrol edilebilir.

* Özellik sayısı fazla ise iki seçenek karşımıza çıkar:

1- Reduce number of features

* Manuel olarak en önemli özellikler hariç kullanmamak tahmin hatası olabilir.

2- Regularization

* Özelliklerin tamamı kalıyor ama parametrelerin değerlerini azaltıyoruz.

Ordinary Diff. Equations:

Derivatives

adi
Ordinary der.
 $\frac{dy}{dt}$

* Bir değişkenli fonksiyonlarda

kısmi
Partial Der.
 $\frac{\partial y}{\partial t}$

* Birden fazla değişkenli fonk'da

Diff. Equations

Ordinary diff.

* Involve one or more ordinary derivatives of unknown func.

Partial Diff

* Involve one or more partial derivatives of unknown func.

Ordinary diff:

$$\frac{dy(t)}{dt} - y(t) = e^t$$

$$\frac{d^2 x(t)}{dt^2} - 5 \frac{dx(t)}{dt} + 2x(t) = \cos(t)$$

unknown func
independent variable

$$\frac{dy}{dt} = 9.8 - \frac{c}{m} y$$

dependent variable

independent var.

order of a diff eq: number of the highest derivative (mer tebe)

$$y'' + xy' - x^3 y = \sin x$$

second order

order of an ordinary diff.

$$\frac{dx(t)}{dt} - x(t) = e^t$$

first order

$$\frac{d^2 x(t)}{dt^2} - 5 \frac{dx(t)}{dt} + 2x(t) = \cos(t)$$

second order

$$\left(\frac{d^2 x(t)}{dt^2} \right)^2 - \frac{dx(t)}{dt} + 2x^4(t) = 1$$

second order

Solution of a diff eq:

* A solution to a diff. eq. is a function that satisfies the eq.

$$\frac{dx(t)}{dt} + x(t) = 0 \quad \text{Solution } x(t) = e^{-t}$$

Proof:

$$\frac{dx(t)}{dt} = -e^{-t} \quad \frac{dx(t)}{dt} + x(t) = -e^{-t} + e^{-t} = 0$$

Linear and Non-linear:

* Bir lineer diff. denk. de tüm türevlerin dereceleri 1 olmalıdır.

* Bağımlı değişken ve türevleri karşın varlinde olmalıdır.

lineerlik formülü

$$a_n(x)y^n + a_{n-1}(x)y^{n-1} + \dots + a_0(x)y = b(x)$$

* eğer a(k)'lerin hepsi sabit ise sabit katsayılı, en az biri sabit değilse değişken katsayılı lineer diff. denk. dir.

* b(x)=0 ise homojen (değilse non-homojen denklemdir).

* f(x)=2x lineer and affine
f(x)=2x+3 non-linear but affine

* Mutlak değer de lineerliği bozar.

* n. mertebeden türevli denk. için eşsiz bir çözüm istiyorsak n tane durumu bilmeniz gerekir.

* Eğer burada verilen durumların hepsi aynı x değerleri içinse initial value, değilse boundary value conditions olur.

$$\ddot{x} + 2\dot{x} + x = e^{-2t}$$

initial
 $x(0)=1, \dot{x}(0)=2.5$

$$\ddot{x} + 2\dot{x} + x = e^{-2t}$$

boundary
 $x(0)=1, x(2)=1.5$
daha zor

* Dif denklemler order, linear, initial-boundary olarak sınıflandırılabilir.

* Analitik çözüm sadece lineer ve bazı özel non-lineer denklemlerde kullanılır.
 * Nümerik çözüm bilinmeyen fonk. un grafiği veya tablosunu elde etmek için kullanılır.

* Bu nümerik yöntemlerin ağırlıklı olarak veya dolaylı olarak kesilmiş Taylor serisi ağırlıklı olarak dayanır.

Separable diff. eq.

* Can be expressed as the product of a function of x and a function of y .

$$\frac{dy}{dx} = g(x) \cdot h(y) \quad h(y) \neq 0 \quad \begin{matrix} dy \cdot y' \text{ leri} = \\ dx \cdot x' \text{ leri} \end{matrix}$$

ÖR: $\frac{dy}{dx} = 2xy^2$ her iki tarafı $\frac{dx}{y^2}$ ile

$$\frac{dy}{y^2} = 2x dx \Rightarrow y^{-2} dy = 2x dx$$

$$\int y^{-2} dy = \int 2x dx \Rightarrow -y^{-1} + C_1 = x^2 + C_2$$

$$-\frac{1}{y} = x^2 + C \Rightarrow y = -\frac{1}{x^2 + C}$$

ÖR: $y' = 6y^2 x$, $y(1) = 1/25$

$$\frac{dy}{dx} = 6y^2 x \Rightarrow \frac{dy}{y^2} = 6x dx$$

$$\int y^{-2} dy = \int 6x dx \Rightarrow -\frac{1}{y} = 3x^2 + C$$

$$-\frac{1}{1/25} = 3 \cdot 1^2 + C \Rightarrow -25 = 3 + C \Rightarrow C = -28$$

$$y = \frac{1}{28 - 3x^2}$$

ÖR: $y' = \frac{3x^2 + 4x - 4}{2y - 4}$, $y(1) = 3$

$$\frac{dy}{dx} = \frac{3x^2 + 4x - 4}{2y - 4} \Rightarrow dy \cdot (2y - 4) = dx \cdot (3x^2 + 4x - 4)$$

$$\int (2y - 4) dy = \int (3x^2 + 4x - 4) dx$$

$$y^2 - 4y = x^3 + 2x^2 - 4x + C$$

$$9 - 12 = 1 + 2 - 4 + C \Rightarrow C = -2$$

$$y^2 - 4y = x^3 + 2x^2 - 4x + 2$$

$$y^2 - 4y + 4 = x^3 + 2x^2 - 4x + 2 + 4$$

$$(y - 2)^2 = x^3 + 2x^2 - 4x + 2$$

$$y - 2 = \pm \sqrt{x^3 + 2x^2 - 4x + 2}$$

$$y = 2 \pm \sqrt{x^3 + 2x^2 - 4x + 2}$$

$$3 = 2 \pm \sqrt{1}$$

$$y = 2 + \sqrt{x^3 + 2x^2 - 4x + 2}$$

ÖR: $y' = \frac{xy^3}{\sqrt{1+x^2}}$, $y(0) = 1$

$$\frac{dy}{dx} = \frac{xy^3}{\sqrt{1+x^2}}$$

$$\int y^{-3} dy = \int \frac{x}{\sqrt{1+x^2}} dx$$

$$-\frac{1}{2y^2} = \sqrt{1+x^2} + C$$

$$-\frac{1}{2} = 1 + C \Rightarrow C = -3/2$$

$$\frac{1}{y^2} = -2\sqrt{1+x^2} + 3 \Rightarrow y^2 = \frac{1}{3 - 2\sqrt{1+x^2}}$$

$$y = \frac{1}{\sqrt{3 - 2\sqrt{1+x^2}}} \quad 1 = \frac{1}{\sqrt{3-2}} = 1$$

$$y = \frac{1}{\sqrt{3 - 2\sqrt{1+x^2}}}$$

Laplace Dönüşümü

$f: [0, \infty) \rightarrow \mathbb{R}$, t deę. baęlı bir fonk. için
 $A \subseteq \mathbb{R}$ $L: A \rightarrow \mathbb{R}$

$$L\{f(t)\} = F(s) = \int_0^{\infty} e^{-st} \cdot f(t) \cdot dt$$

NOT: Dönüşümün iyi tanımlı olabilmesi için

$\int_0^{\infty} e^{-st} f(t) dt$ has olmayan int. yakınsak olması gerekir, yani:

$$\int_0^{\infty} e^{-st} f(t) dt = \lim_{R \rightarrow \infty} \int_0^R e^{-st} f(t) dt < \infty$$

NOT: Bu integral hesaplanırken s sabit olarak düşünülür ancak durumdaki fonk. un deęiskenidir.

Laplace Tablosu

$$L\{1\} = 1/s \quad (s > 0) \quad L\{a\} = a/s \quad (s > 0)$$

$$L\{t^n\} = n! / s^{n+1}, \quad n = 1, 2, 3, \dots \quad (s > 0)$$

$$L\{e^{at}\} = 1/s - a \quad (s > a)$$

$$L\{\sin kt\} = k / s^2 + k^2 \quad (s > 0)$$

$$L\{\cos kt\} = s / s^2 + k^2 \quad (s > 0)$$

$$L\{\sinh kt\} = k / s^2 - k^2$$

$$L\{\cosh kt\} = s / s^2 - k^2$$

Laplace Dönüşümünün Liniyerlięi

$$L\{c_1 f(t) + c_2 g(t) - c_3\} = c_1 \cdot L\{f(t)\} + c_2 \cdot L\{g(t)\} - c_3$$

s 'de öteleme

$$L\{f(t)\} = F(s) \quad (s > s_0) \Rightarrow L\{e^{at} f(t)\} = F(s-a)$$

* Önce $f(t)$ 'nin Laplace'ini hesapla, bundan sonra $F(s)$ 'deki s yerine $s-a$ yaz.

$$L\{e^{at} \cdot b\} = \frac{b}{s-a}$$

$$L\{e^{at} \cdot t^n\} = \frac{n!}{(s-a)^{n+1}}$$

$$L\{e^{at} \cdot \sin(bt)\} = \frac{b}{(s-a)^2 + b^2}$$

$$L\{e^{at} \cdot \cos(bt)\} = \frac{s-a}{(s-a)^2 + b^2}$$

$$L\{e^{at} \cdot e^{bt}\} = L\{e^{t(a+b)}\} = \frac{1}{s-(a+b)}$$

Laplace Dönüşümünün Türevleri

$$L\{f(t)\} = F(s) \text{ olsun}$$

$$L\{t^n \cdot f(t)\} = (-1)^n \cdot \frac{d^n}{ds^n} \{F(s)\}$$

NOT: $L\{f(t)\} = F(s)$ ve $\lim_{t \rightarrow 0^+} \frac{f(t)}{t}$

mevcut ise:

$$L\{1/t \cdot f(t)\} = \int_s^{\infty} F(m) dm$$

Türev Fonksiyonunun Laplace Dönüşümü

* f ve f' paralı süreklili ve α mertebesindeki üstel olsunlar. $s > \alpha$ için:

$$L\{f'(t)\} = s \cdot L\{f(t)\} - f(0)$$

* Genel olarak $n \in \mathbb{N}$ için

$$L\{f^{(n)}(t)\} = s^n \cdot L\{f(t)\} - s^{n-1} f(0) - s^{n-2} f'(0) - \dots - s f^{(n-2)}(0) - f^{(n-1)}(0)$$

Tam olmayan üstel Laplace

$$L\{t^{n-1/2}\} = \frac{1 \cdot 3 \cdot 5 \cdot \dots \cdot (n-1) \sqrt{\pi}}{2^n \cdot s^{n+1/2}} \quad (n=1, 2, 3, \dots)$$

Ters Laplace Dönüşümü

* $\mathcal{L}\{f(t)\} = F(s)$, $f: [0, \infty) \rightarrow \mathbb{R}$ sürekli fonk. için Laplace dönüşümünün tersi

$$\mathcal{L}^{-1}\{F(s)\} = f(t)$$

* İki fonksiyon aynı Laplace dönüşümüne sahipse bu fonk. lar mutlaka birbirine eşittir. (süreklilerse)

Ters Laplace Dönüşümünün Lineerliği

$$\mathcal{L}^{-1}\{a \cdot F(s) + b \cdot G(s)\} = a \mathcal{L}^{-1}\{F(s)\} + b \mathcal{L}^{-1}\{G(s)\}$$

İntegral Laplace Dönüşümü

$$\mathcal{L}\{f(t)\} = F(s), t > 0, s > 0, s > \infty$$

$$\mathcal{L}\left\{\int_0^t f(t) dt\right\} = 1/s F(s)$$

$$\int_0^t f(t) dt = \mathcal{L}^{-1}\{1/s F(s)\}$$