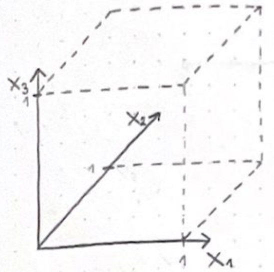
SU1: 1. DOMAĆA ZADAĆA

V02 - zadaci s ispita

3. $X = \{0,1\}^3 \qquad \theta \in \mathbb{R}^6$



$$\Theta = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \\ \theta_{31} & \theta_{32} \end{bmatrix}$$

ekstenzionalna jednakost: $h_1 = h_2 \leftrightarrow \forall x\,(h_1(x) = h_2(x)),\ x \in X$

- parametri oblikuju kvadar - ako se točka nalazi unutar kvadra onda je $y=1$, inače 0
  - ↳ razlikujemo slučajeve kada kvadar 'obuhvaća':
    1) jednu točku: 8
    2) jedan brid: 12
    3) jednu plohu: 6
    4) sve točke: 1
    5) niti jednu točku

$$\left.\begin{array}{l} 8 \\ 12 \\ 6 \\ 1 \\ \end{array}\right\} |H| = 28$$

7. Asimetrična funkcija gubitka kod detekcije karcinoma treba više kažnjavati lažno negativne od lažno pozitivnih primjera: $L(1,0) > L(0,1)$ i $L(1,1) = L(0,0) = 0$

V03 - zadaci s ispita

3. $h(x; w_0, w_1) = w_0 + w_1 x \qquad x \in \mathcal{N}(-1+2x, \sigma^2)$

$L(y, h(x)) = (y + h(x))^2$

- zbog malo šuma možemo gledati kao da se primjeri ravnaju po funkciji $-1+2x$, što znači da uvrštavamo tu funkciju bez šuma u funkciju gubitka:

$$L(y, h(x)) = (y + h(x))^2 = (-1 + 2x + w_0 + w_1 x)^2 = \left[1(-1 + w_0) + x(2 + w_1)\right]^2$$

- da bi gornji izraz bio minimalan, mora biti $w_0 = 1$ i $w_1 = -2$

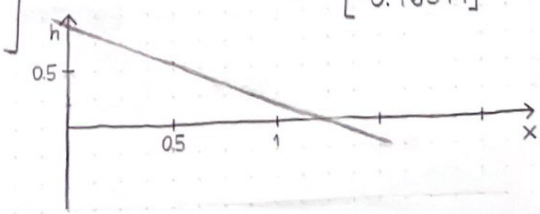
V04 - zadaci za učenje

3. $D = \{(0.25, 0.707), (0.5, 1), (1, 0), (1.5, -1), (2, 0)\} \qquad f(x) = \sin(\pi x)$

a) $\phi(x) = (1, x)$

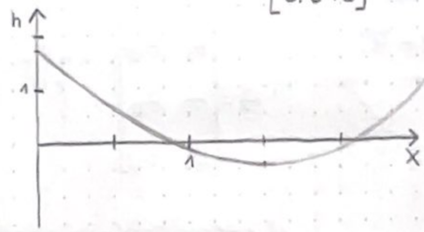
$$\Phi = \begin{bmatrix} 1 & 0.25 \\ 1 & 0.5 \\ 1 & 1 \\ 1 & 1.5 \\ 1 & 2 \end{bmatrix} \qquad y = \begin{bmatrix} 0.707 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \qquad \vec{w} = (\Phi^T \Phi)^{-1} \Phi^T y = \begin{bmatrix} 0.94329 \\ -0.76371 \end{bmatrix}$$

$h(x) = 0.94 - 0.76x$



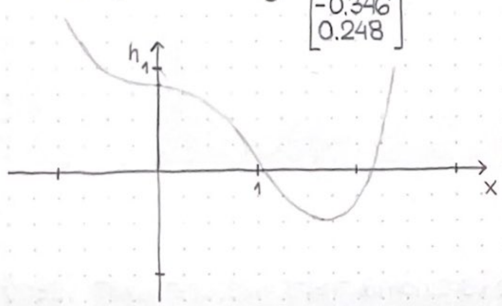
b) $\Phi = \begin{bmatrix} 1 & 0.25 & 0.0625 \\ 1 & 0.5 & 0.25 \\ 1 & 1 & 1 \\ 1 & 1.5 & 2.25 \\ 1 & 2 & 4 \end{bmatrix}$ $\vec{y} = \begin{bmatrix} 0.707 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$ $\vec{w} = (\Phi^T\Phi)^{-1}\Phi^T y = \begin{bmatrix} 1.754 \\ -2.941 \\ 0.976 \end{bmatrix}$

$h(x) = 1.75 - 2.94x + 0.98x^2$



c) $\Phi = \begin{bmatrix} 1 & 0.25 & 0.0625 & 0.015625 & 0.00390625 \\ 1 & 0.5 & 0.25 & 0.125 & 0.0625 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1.5 & 2.25 & 3.375 & 5.0625 \\ 1 & 2 & 4 & 8 & 16 \end{bmatrix}$ $y = \begin{bmatrix} 0.707 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$

$w = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T y = \begin{bmatrix} 0.833 \\ -0.282 \\ -0.415 \\ -0.346 \\ 0.248 \end{bmatrix}$ $h(x) = 0.83 - 0.28x - 0.41x^2 - 0.35x^3 + 0.25x^4$

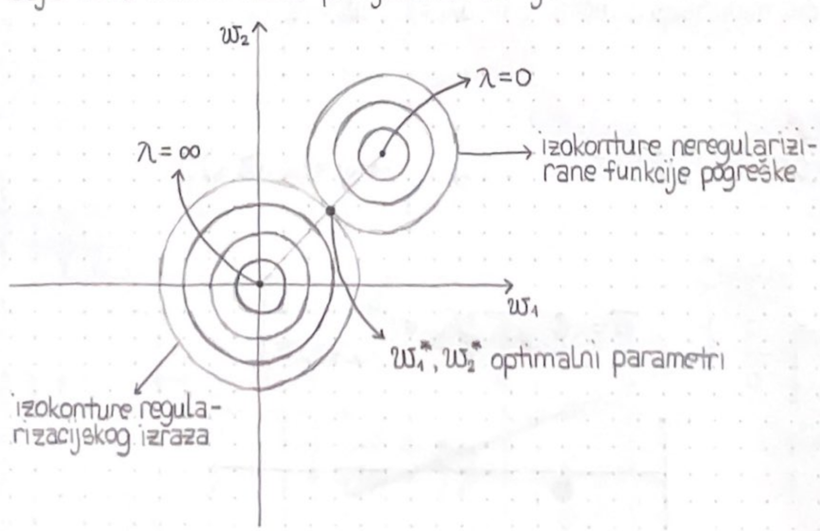


d) Najprikladniji je model pod c). Njegov graf čak i izgledom najviše podsjeća na sinusoidu, a ima i najmanju kvadratnu pogrešku.

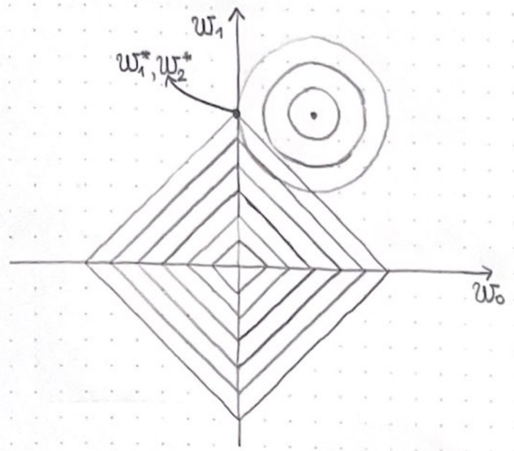
5. a) Svrha regularizacije jest sprečavanje prenaučenosti modela ograničavanjem rasta vrijednosti parametara. Temelji se na pretpostavci da, što je model složeniji, to su mu veće vrijednosti parametara.

b) Glavna prednost regulariziranog modela u odnosu na neregularizirani jest ta da je regularizirani model teže prenaučiti, što pogotovo dolazi do izražaja kada imamo malo primjera za učenje.

c)



d)



Kod L1-regularizacije će se zbog oštrijih izokontura regularizacijskog izraza lakše dogoditi da se minimizator regularizirane funkcije pogreške nađe na nekoj od koordinatnih osi prostora parametara. Ukoliko se to dogodi, druga težina priteže se na nulu, što rezultira rjeđim modelima.

V04 - Zadaci s ispita

5. Hipoteza $H_{2,0}$ imat će veću pogrešku na skupu za učenje od hipoteze $H_{5,0}$ zato što će se zbog manje vrijednosti hiperparametra d moći manje prilagoditi podacima za učenje. S druge strane, ne znamo koji d je optimalan, stoga ove dvije hipoteze mogu podjednako loše generalizirati.