~~Starting~~ point:

Terminology:

$\mathcal{X}$ ~~$\mathcal{Y}$~~

~~input~~

$\mathcal{X}$: input space

$\mathcal{Y}$: set of target values

For simplicity:

$$\mathcal{Y} = \{0,1\}$$

Concept: concept class

$c: \mathcal{X} \to \mathcal{Y}$ $\quad c \in C$ $\quad \to$ ~~Samples~~ Samples i.i.d. w.r.t. D on $\mathcal{X}$

we want to learn this

$S = (x_1, \dots, x_m)$

↳ labels $(c(x_1), \dots, c(x_m))$

Hypothesis class

Task: Given ~~...~~ S, find suitable $h_S \in H$

~~Hypothesis:~~

Selection: Error function:

For $h \in H$ and concept "c", underlying distribution D:

Generalization Error / Risk:

$$R(h) = \Pr_{x \sim D}[h(x) \neq c(x)] = \mathop{E}_{x \sim D}[1_{h(x) = c(x)}]$$

More accesible :

Empirical Error / Empirical Risk:

For $h \in H$ and target concept "c", ~~for~~ sample $S = (x_1, \dots, x_m)$:

$$\hat{R}(h) = \frac{1}{m} \sum_{i=1}^{m} 1_{h(x_i) \neq c(x_i)}$$

PaC - Frame work :

Concept class C is (by. def.) PAC learnable iff: $\exists$ algorithm A, poly $(\cdot,\cdot,\cdot,\cdot)$ s.t. $\forall \varepsilon > 0$, $\delta > 0$ $\forall D$ on $\mathcal{X}$ $\forall c \in C$:

$$\Pr_{S \sim D^m} [R(h_S) \leq \varepsilon] \geq 1 - \delta$$

for $m \geq \text{poly}(1/\varepsilon, 1/\delta, n, \text{size } c)$

Last time: Bounds on $R(h_S)$!

Consistant case ($\hat{R}(h_S) = 0$) :

with probability at least $(1-\delta)$,

$R(h_S) \leq \frac{1}{m}(\log|H| + \log\frac{1}{\delta})$ $\Leftarrow\Rightarrow$ $\forall \varepsilon, \delta > 0$: $\Pr_{S \sim D^m}[R(h_S) \leq \varepsilon] \geq 1 - \delta$ holds if $m \geq \frac{1}{\varepsilon}(\log|H| + \log\frac{1}{\delta})$

Inconsistant case:

$|H| < \infty$, $\forall \delta > 0$ with prob. at least $1-\delta$:

$$\forall h \in H, \ R(h) \leq \hat{R}(h) + \sqrt{\frac{\log|H| + \log\frac{2}{\delta}}{2m}}$$

Problem: $|H| \overset{!}{<} \infty$

Remedy today!

First: Introduce different measure of complexity!

Definition: Empirical Rademacher complexity

$G \ni g : Z \to [a,b]$, $S = (z_1, \dots, z_m)$ a fixed sample of size m.
$\in Z^m$

$$\hat{R}_S(G) := \underset{\sigma}{E}\left[\sup_{g \in G} \frac{1}{m} \sum_{i=1}^{m} \sigma_i g(z_i)\right],$$

where $\sigma = (\sigma_1, \dots, \sigma_m)^T$ with the σ_i's i.i.d. (uniformly), taking values in $\{-1, 1\}$. "σ_i's" are called Rademacher variables.

what is happening?

$g_S := (g(z_1), \dots, g(z_m))^T$

$\sigma \cdot g_S$: measures correlation of g_S with random noise

$\sup_{g \in G} \frac{\sigma \cdot g_S}{m}$: measures how well G correlates with σ over S.

$\hat{R}_S$: measures on average how well "G" correlates with random noise

Definition: Rademacher complexity.

Let samples be drawn according to D. Define:

$$\mathcal{R}_m(G) := \underset{S \sim D^m}{E}[\hat{\mathcal{R}}_S(G)]$$

We will need:

McDiarmids inequality:

Let $(Z_1, \dots, Z_N) = Z$ be a finite seq. of independent random variables, each with values in $\mathcal{Z}$ and $\varphi: \mathcal{Z}^n \to \mathbb{R}$ a meas. fct. s.t. $|\varphi(z) - \varphi(z')| \leq \nu_i$ whenever z and z' only differ in the ith coordinate. Then $\forall \varepsilon > 0$

$$\mathbb{P}[\varphi(Z) - E[\varphi(Z)] \geq \varepsilon] \leq \exp\left[-\frac{2\varepsilon^2}{\sum_{i=1}^{n} \nu_i}\right]$$

(same for ~~...~~ $\varphi \leftrightarrow E(\varphi)$).

Theorem:

Let G be a family of functions mapping from Z to $[0,1]$. Then $\forall \delta > 0$ with probability at least $1-\delta$:

$\forall g \in G$:

$$E[g(z)] \leq \frac{1}{m}\sum_{i=1}^{m} g(z_i) + 2\mathcal{R}_m(G) + \sqrt{\frac{\log\frac{1}{\delta}}{2m}} \quad (\mathrm{I})$$

$$E[g(z)] \leq \frac{1}{m}\sum_{i=1}^{m} g(z_i) + 2\hat{\mathcal{R}}_S(G) + 3\sqrt{\frac{\log\frac{2}{\delta}}{2m}}$$

Pf.: For $S = (z_1, \dots, z_m)$, $g \in G$, ~~A~~ Let:

$\hat{E}_S[g] := \sum_{i=1}^{m} g(z_i)$.

Define

$$\phi(S) := \sup_{g \in G} E[g] - \hat{E}_S[g].$$

Let S and S' be two samples differing in exactly one point, e.g. z_m in S and z'_m in S'.

Then

$$\phi(S') - \phi(S) \leq \sup_{g \in G} \left[\hat{E}_S[g] - E_{S'}[g] \right]$$

$$= \sup_{g \in G} \frac{g(z_m) - g(z'_m)}{m} \leq \frac{1}{m}$$

$g: Z \to [0,1]$.

Similarly

$\phi(S) - \phi(S') \leq \frac{1}{m}$. ~~Th~~ Thus $|\phi(S) - \phi(S')| \leq 1/m$

Mc Diarmid, $c_i = c/m$:

$$\mathbb{P}[\phi(S) - \mathbb{E}_S[\phi(S)] \geq \varepsilon] \leq e^{-2m\varepsilon^2} =: \delta$$

→ Solving for ε:

$$\phi(S) \leq \mathop{E}_{S}[\phi(S)] + \sqrt{\frac{\log \frac{1}{\delta}}{2m}}$$

Next we bound

$$\mathop{E}_{S}[\phi(S)] = \mathop{E}_{S}[\sup_{g\in H} E[g] - \hat{E}_S(g)]$$

$$= \mathop{E}_{S}[\sup_{g\in H} \mathop{E}_{S'}[\hat{E}_{S'}(g) - \hat{E}_S(g)]]$$

$$\leq \mathop{E}_{S,S'}[\sup_{g\in H}(\hat{E}_{S'}(g) - \hat{E}_S(g))]$$

Jenssen; $\sup(\cdot)$ is convex

Convex

$$\varphi\left(\int_\Omega g\, d\mu\right) \leq \int_\Omega \varphi\circ g\, d\mu$$

$$= \mathop{E}_{S,S'}\left[\sup_{g\in H} \frac{1}{m}\sum_{i=1}^{m}\left(g(z_i') - g(z_i)\right)\right]$$

$$= \mathop{E}_{\sigma,S,S'}\left[\sup_{g\in H} \frac{1}{m}\sum_{i=1}^{m} \sigma_i(g(z_i') - g(z_i))\right]$$

if $\sigma_i = 1$: no problem

if $\sigma_i = -1$: switch $z_i \longleftrightarrow z_i'$ in S and S'.

But we average over S, S'.

$$\leq \mathop{E}_{\sigma,S'}\left[\sup_{g\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i g(z_i)\right] + \mathop{E}_{\sigma,S}\left[\sup_{g\in H}\frac{1}{m}\sum_{i=1}^{m} -\sigma_i g(z_i)\right]$$

$$= 2\mathop{E}_{\sigma,S}\left[\sup_{g\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i g(z_i)\right] = 2\mathfrak{R}_m(G).$$

Thus

(I): $E[g(z)] \cancel{\leq} \cancel{\frac{1}{m}} \cancel{\sum_{i=1}^{m}}$

$$\phi(S) \leq 2 \mathcal{R}_m + \sqrt{\frac{\log \frac{1}{\delta}}{2m}}$$

$$\Updownarrow \quad \phi(S) := \sup_{g \in G} E(g) - \hat{E}_S(g)$$

With prob. at least $1-\delta$

$$E(g(z)) \leq \sup_{g \in G} E(g(z)) \leq \hat{E}_S[g] + 2 \mathcal{R}_m + \sqrt{\frac{\log \frac{1}{\delta}}{2m}}$$

For II: $\delta \to \delta/2$:

Then with prob. at least $1-\delta/2$:

$$E(g(z)) \leq \hat{E}_S[g] + 2 \hat{\mathcal{R}}_S(G) + \sqrt{\log \tfrac{2}{\delta} / 2m}$$

Note:

$\cancel{\hat{R}_{S'}(G) - \hat{R}_{S'}(G) = E_\sigma (\sup_{g \in G} g(z)} \; E$

$$\hat{R}_{S'}(G) - \hat{R}_S(G) = E_{\sigma'}\left[\sup_{g \in G} \frac{1}{m}\left(\sum_{i=1}^{m-1} \sigma_i' g(z_i') + \sigma_m' g(z_m')\right)\right]$$

$$- E_{\sigma}\left[\sup_{g \in G} \frac{1}{m} \sum_{i=1}^{m} g(z_i)\, \sigma_i\right]$$

$$\leq \frac{1}{m}$$

$\sigma_m' g(z_m') \leq \sigma_m' g(z_m) + 1$

→ McDiarmid again:

$$\mathfrak{R}_m \leq \hat{\mathfrak{R}}_S(G) + \sqrt{\frac{\log \frac{2}{\delta}}{2m}} \quad \text{with prob. at least } 1-\delta/2$$

→ with probability at least $1-\delta$

$$E(g(z)) \leq \hat{E}_S[g] + 2\hat{\mathfrak{R}}_S(G) + 3\sqrt{\frac{\log \frac{2}{\delta}}{2m}}$$

//

Lemma:

Let H be a family of functions taking values in $\{-1,+1\}$.
Let G be the family of zero-one loss fcts. associated to G:

$$G = \{ (x,y) \mapsto 1_{h(x) \neq y} : h \in H \}.$$

For any Sample $S=((x_1,y_1),\dots,(x_m,y_m))$ of elements in $\mathcal{X} \times \{-1,1\}$, Let $S_{\mathcal{X}} = (x_1,\dots,x_m)$. Then

$$\hat{\mathfrak{R}}_S(G) = \frac{1}{2}\hat{\mathfrak{R}}_{S_{\mathcal{X}}}(H).$$

Pf.:

$$\hat{R}_S(G) = \underset{\sigma}{E}\left[\sup_{h\in H} \frac{1}{m}\sum_{i=1}^{m} \sigma_i 1_{h(x_i)=y_i}\right]$$

$$= \underset{\sigma}{E}\left[\sup_{h\in H} \frac{1}{m}\sum \sigma_i \frac{1-y_i h(x_i)}{2}\right]$$

$$= \frac{1}{2}\underset{\sigma}{E}\left[\sup_{h\in H} \frac{1}{m}\sum_{i=1}^{m} -\sigma_i y_i h(x_i)\right]$$

$$= \frac{1}{2}\underset{\sigma}{E}\left[\sup_{h\in H} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h(x_i)\right] = \frac{1}{2}\mathfrak{R}_{S_{\mathcal{X}}}(H).$$

Note: this implies $\mathfrak{R}_m(G) = \frac{1}{2}\mathfrak{R}_m(H)$ by taking expectations!

→ use this for bounds on Risk function!

Theorem: Rademacher Complexity bounds (binary)

Let H be a family of functions, taking values in $\{-1, +1\}$.

Let D be the dist. over the input space $\mathcal{X}$.

Then $\forall \delta > 0$ with probability atleast $1-\delta$ over ~~a sample~~ a sample S. $\forall h \in H$:

$$R(h) \leq \hat{R}(h) + \mathfrak{R}_m(H) + \sqrt{\frac{\log \frac{1}{\delta}}{2m}} \qquad (\text{III})$$

$$R(h) \leq \hat{R}(h) + \hat{\mathfrak{R}}_S(H) + 3\sqrt{\frac{\log \frac{2}{\delta}}{2m}} \qquad (\text{IV})$$

Pf. Apply the two previous results!

(IV) is Data dependent

Can we compute $\hat{\mathfrak{R}}_S(H)$? Note

$$\hat{\mathfrak{R}}_S(H) = \underset{\sigma}{E}\left[\sup_{h\in H} \frac{1}{m}\sum_{i=1}^{m} -\sigma_i h(x_i)\right] = -\underset{\sigma}{E}\left[\inf_{h \in H} \frac{1}{m} \sigma_i h(x_i)\right]$$

For fixed value of σ, computing $\inf_{h\in H} \frac{1}{m}\sum_{i=1}^{m} \sigma_i h(x_i)$ is empirical risk minimization → could be computationally hard. → Better estimates?

Growth function:

Def.: Growth function:

$\Pi_{H}: \mathbb{N} \to \mathbb{N}$ for a hypothesis set H is defined via:

$$\forall m \in \mathbb{N}: \Pi_{H}(m) = \max_{\{x_1,\dots,x_m\} \subseteq X} \left| \left\{ \left(h(x_1), \dots h(x_m)\right) : h \in H \right\} \right|$$

Maximum # of distinct ways in which m points can be classified.

Lemma: Growth function under compositions:

Consider function classes:

$\mathcal{F}_1 \subseteq \mathcal{Y}^{\mathcal{X}}$, $\mathcal{F}_2 \subseteq \mathcal{Z}^{\mathcal{Y}}$ and $\mathcal{F} = \mathcal{F}_2 \circ \mathcal{F}_1$:

The respective growth functions satisfy

$\Gamma(n) \leq \Gamma_1(n)\, \Gamma_2(n)$

Proof:

Fix an arbitrary subset $A \subseteq \mathcal{X}$ of cardinality n.

With $\mathcal{G} := \mathcal{F}_1|_A$, we can write:

$$\mathcal{F}|_A = \bigcup_{g \in \mathcal{G}} \{ f \circ g \mid f \in \mathcal{F}_2 \}.$$ Thus:

$$|\mathcal{F}|_A| \leq |\mathcal{F}_1|_A| \max_{g \in \mathcal{G}} |\{ f \circ g \mid f \in \mathcal{F}_2 \}|$$

$$\leq \Gamma_1(n) \max_{g \in \mathcal{G}} |\mathcal{F}_2|_{g(A)}|$$

$$\leq \Gamma_1(n)\, \Gamma_2(n)$$

//

Next we want to relate the growth Function to the Rademacher complexity.

We use Massart's lemma:

Lemma:

Let $A \subseteq \mathbb{R}^m$ be a finite set with $r = \max_{x \in A} \|x\|_2$.

Then:

$$\mathop{E}_{\sigma}\left[\frac{1}{m} \sup_{x \in A} \sum_{i=1}^{m} \sigma_i x_i\right] \leq \frac{r\sqrt{2\log|A|}}{m},$$

where σ_i's are independent uniform random variables, taking values in $\{-1,+1\}$ and $x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$.

Pf.: :

For $\forall t>0$:

$$\exp\left(t \mathop{E}_{\sigma}\left[\sup_{x\in A} \sum_{i=1}^{m} \sigma_i x_i\right]\right) \overset{\text{Jenssen}}{\leq} E_\sigma \exp\left[t \sup_{x\in A} \sum_{i=1}^{m} \sigma_i x_i\right])$$

$$= \mathop{E}_{\sigma} \sup_{x\in A} \exp\left[t \sum_{i=1}^{m} \sigma_i x_i\right]) \leq \sum_{x\in A} \mathop{E}_{\sigma}\left(\exp\left[t \sum_{i=1}^{m} \sigma_i x_i\right]\right).$$

Since σ_i's are independent

$$\sum_{x\in A} \mathop{E}_{\sigma} \exp\left(t \cancel{\sup_{x\in}} \sum_{i=1}^{m} \sigma_i x_i\right) \leq \sum_{x\in A} \prod_{i=1}^{m} \mathop{E}_{\sigma_i} \exp[t\sigma_i x_i]$$

$$\underset{\text{Hoeffding}}{\leq} \sum_{x\in A} \prod_{i=1}^{m} \exp\left[\frac{t^2 (2x_i)^2}{8}\right]$$

$$= \sum_{x\in A} \exp\left[\frac{t^2}{2} \sum_{i=1}^{m} x_i^2\right] \leq \sum_{x\in A} \exp\left[\frac{t^2 r^2}{2}\right] = |A| e^{\frac{t^2 r^2}{2}}$$

Hoeffding: X be any R.V. with $E[X]=0$ and s.t. $a \leq X \leq b$. The $\forall \lambda \in \mathbb{R}$, $E[e^{\lambda X}] \leq \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right)$

$a \leq 0 \leq b$!

log and divide by t:

$$\mathop{E}_{\sigma}\left[\sup_{x\in A} \sum_{i=1}^{m} \sigma_i x_i\right] \leq \frac{\log|A|}{t} + \frac{t r^2}{2}$$

minimize upper bound at $t = \frac{\sqrt{2 \log |A|}}{r}$:

$$\mathop{E}_{\sigma}\left[\sup_{x\in A} \sum_{i=1}^{m} \sigma_i x_i\right] \leq r \sqrt{2 \log |A|}$$

Corr.:

Let G be a family of fcts. $\to \{-1, +1\}$. Then:

$$\mathcal{R}_m(G) \leq \sqrt{\frac{2 \log \Gamma_G(m)}{m}}$$

Proof:

For a fixed sample $S = (x_1, \dots, x_m)$, we denote by $G|_S$ the set of vectors of function values $(g(x_1), \dots, g(x_m))^T$ where $g \in G$. Since values are in $\{-1, +1\}$, $\sqrt{m}$ is norm bound. Then apply Massarts lemma as follows

$$R_m(G) = \mathop{E}_S \left[\mathop{E}_\sigma \left[\sup_{u \in G|_S} \frac{1}{m} \sum_{i=1}^m \sigma_i u_i \right] \right] \leq \mathop{E}_S \left[\frac{\sqrt{m} \sqrt{2 \log |G|_S|}}{m} \right]$$

By def. $|G|_S| \leq \Gamma_G(m)$. Thus

$$\mathcal{R}_m(G) \leq \mathop{E}_S \left[\frac{\sqrt{m} \sqrt{2 \log \Gamma_G(m)}}{m} \right] = \sqrt{\frac{2 \log \Gamma_G(m)}{m}}$$

Cor.: Growth function generalization bound:

H be a family of functions taking values in $\{-1, +1\}$.
Then $\forall \delta > 0$ with prob. at least $1-\delta$, $\forall h \in H$:

$$R(h) \leq \hat{R}(h) + \sqrt{\frac{2 \log \Pi_H(m)}{m}} + \sqrt{\log\left(\frac{1}{\delta}\right)/2m}$$

Growth function eval. for ~~than~~ every m tedious.

$\Rightarrow$ Scalar measure of complexity?

Vapnik - Chervonenkis dimension

If $|\mathcal{Y}| = 2$: $\Gamma(n)$ behavior peculiar!

$\exists\, d \in \mathbb{N}_0^{\infty}$ s.t. $\Gamma(n) = 2^n$ for $n \leq d$

$\Gamma(n)$ poly. bdd. for $n > d$.

Def.:

For $\mathcal{F} \subseteq \mathcal{Y}^{\mathcal{X}}$ with $\mathcal{Y}$ binary:

$$VCdim(\mathcal{F}) := \max\{n \in \mathbb{N}_0 \mid \Gamma(n) = 2^n\}$$

if the max. exists and $VCdim(\mathcal{F}) = \infty$ otherwise

In other words

If $VCdim(\mathcal{F}) = d$, then $\exists\, A \subseteq \mathcal{X}$ of d pts.

s.t. $\mathcal{F}|_A = \mathcal{Y}^A$ and the VC dim is the largest

such number #

Example I:

If $\mathcal{F} = \{ \mathbb{R} \ni x \mapsto \operatorname{sgn}[x-b] \}_{b \in \mathbb{R}}$, then

VC dim $(\mathcal{F}) = 1$ since $\Gamma(n) = n+1$

Example II:

Lem: Can prove:

Let $\mathcal{G}$ be a real vector space of functions from $\mathcal{X}$ to $\mathbb{R}$ and

$\phi \in \mathbb{R}^{\mathcal{X}}$. Then $\mathcal{F} := \{ x \mapsto \operatorname{sgn}[g(x) + \phi(x)] \}_{g \in \mathcal{G}} \subseteq \{-1, +1\}^{\mathcal{X}}$

has VC dim $(\mathcal{F}) = \dim(\mathcal{G})$

example:

If $\mathcal{F} = \{\mathbb{R} \ni x \mapsto \operatorname{sgn}[x-b]\}_{b \in \mathbb{R}}$, then

VC dim $(\mathcal{F}) = 1$ since $\Gamma(n) = n+1$.

Theorem: VC-dichotomy of growth function

Consider a function class $\mathcal{F} \subseteq \mathcal{Y}^{\mathcal{X}}$ with binary target space $\mathcal{Y}$ and VC-dim d. Then

$$\Gamma(n) = \begin{cases} 2^n & \text{if } n \le d \\ \le \left(\frac{en}{d}\right)^d & \text{if } n > d \end{cases}$$

Pf.: $\Gamma(n) = 2^n \;\; \forall n \le d$ holds by def. of the VC dim.

For $n > d$, we show that $\forall A \subseteq \mathcal{X}$ with $|A| = n$, the following is true:

$$|\mathcal{F}|_A| \le |\{B \subseteq A \mid \mathcal{F}|_B = \mathcal{Y}^B\}|.$$

If this holds, we can upper bound the r.h.s. by

$$\sum_{i=0}^{d} \binom{n}{i} \le \sum_{i=0}^{n} \binom{n}{i} \left(\frac{n}{d}\right)^{d-i}$$

$$= \left(\frac{n}{d}\right)^d \left(1 + \frac{d}{n}\right)^n \le \left(\frac{n}{d}\right)^d e^d.$$

$\square$

Hence: prove

$$|\mathcal{F}|_A| \leq |\{B \subseteq A \mid \mathcal{F}|_B = Y^B\}|$$

Induction on $|A|$.

For $|A| = 1$: true (note that $\emptyset \subseteq B$)

Now assume it holds for $|A| = n-1$.

Let a be any element of A and define

$$\mathcal{F}' := \{h \in \mathcal{F}|_A \mid \exists g \in \mathcal{F}|_A : h(a) = g(a) \wedge (h-g)|_{A \setminus a} = 0\}$$

$$\mathcal{F}_a := \mathcal{F}'|_{A \setminus a}.$$

Then $|\mathcal{F}_A| = |\mathcal{F}_{A \setminus a}| + |\mathcal{F}_a|$. Both terms on r.h.s. can be bdd. by induction hypothesis.

For the first term:

$$|\mathcal{F}|_{A \setminus a}| \leq |\{B \subseteq A \mid \mathcal{F}|_B = Y^B \wedge a \notin B\}|$$

The second term can be bdd. by

$$|\mathcal{F}_a| = |\mathcal{F}'|_{A \setminus a}| \leq |\{B \subseteq A \setminus a \mid \mathcal{F}'|_B = Y^B\}|$$

$$= |\{B \subseteq A \setminus a \mid \mathcal{F}'|_{B \cup a} = Y^{B \cup a}\}|$$

$$= |\{B \subseteq A \mid \mathcal{F}'|_{B} = Y^B \wedge a \in B\}|$$

$$\leq |\{B \subseteq A \mid \mathcal{F}_B = Y^B \wedge a \in B\}|$$

Cor :::

VC Dim gen. bounds

Let H be a family of functions taking values in $\{-1, +1\}$ with VC dim d.

Then $\forall \delta > 0$ with prob. at least $1-\delta$, the following holds $\forall h \in H$:

$$R(h) \leq \hat{R}(h) + \sqrt{\frac{2d \log \frac{em}{d}}{m}} + \log \sqrt{\frac{\log \frac{1}{\delta}}{2m}}$$

$$\rightarrow R(h) \leq \hat{R}(h) + \mathcal{O}\left(\sqrt{\frac{\log m/d}{m/d}}\right)$$