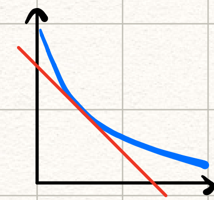
Convexity : Central role in local variational framework

Example : Approx. $f(x) = \exp(-x)$ by lower bound $g(x, \lambda)$

the "variable"

Convex function of x



BOUNDED

Goal : $\boxed{\text{Approx. } f(x)}$ by simpler linear function $g(x)$

POINT : $g(x)$ is approximating the function $f(x)$

- Approximation $g(x)$ is a functional being optimized

Tangent line using $\boxed{\text{first order Taylor expansion}}$ at $x = \xi$

Tangent line of convex func. bellow func.

$$g(x) = f(\xi) + f'(\xi)(x - \xi) \leq f(x)$$

Variational parameter ξ of functional $g()$

Noting that $f(\xi) = \exp(-\xi)$ 　　Property of function $f(x)$ being approximated

$f'(\xi) = -\exp(-\xi)$

$$\Rightarrow \quad g(x) = \exp(-\xi) - \exp(-\xi)(x - \xi) \quad \text{Taylor expansion}$$

Substitute ξ by λ

Reparametrization $\underbrace{-\exp(-\xi)}_{h(\xi)} = \lambda$ 　$| -\log$

$\xi = \underbrace{-\log(-\lambda)}_{h^{-1}(\lambda)}$

1. Start with a common component $h(\xi) = \lambda$
2. Find inverse function $\xi = h^{-1}(\lambda)$

$$\Rightarrow \; t(x) = -\lambda + \lambda\,(x + \log(-\lambda))$$
$$= \lambda x - \lambda + \lambda \log(-\lambda)$$

Approx. expressed in reparameterization λ

Best approximation : $\boxed{\text{Maximizing the lower bound}}$ $\;t(x) \le f(x)$

λ^* depends on x!

$$\Rightarrow \; f(x) = t(x, \lambda^*) \ge t(x, \lambda)$$

for $x = \xi$ (at this point)

$$t(x, \lambda^*) = \max_{\lambda} \left[\lambda x - \lambda + \lambda \log(-\lambda)\right]$$

Variational parameter

Optimize for λ to obtain tightest bound

General formulation : Framework of convex duality

Goal : Approx. convex function $f(x)$ by

- Linear function λx
- Intercept $g(\lambda) = -\min_x (f(x) - \lambda x) = \max_x (\lambda x - f(x))$

$$\Rightarrow \; \boxed{t(x, \lambda) = \lambda x - g(z)}$$

$$\Rightarrow \; f(x) = t(x, \lambda^*) = \max_{\lambda} \left[\lambda x - t(\lambda)\right]$$

POINT : Dual role of $f(x)$, $g(\lambda)$

$$\boxed{\begin{aligned} f(x) &= \max_{\lambda} \left[\lambda x - t(\lambda)\right] \\ g(\lambda) &= \max_{x} \left[\lambda x - f(x)\right] \end{aligned}}$$

Expressing Approx = True func. In both perspectives

Applying duality relations to obtain approx. for $f(x) = e^{-x}$

$$g(\lambda) = \max_x \left[\lambda x - f(x) \right] = \max_x \left[\lambda x - \exp(-x) \right]$$

$$\frac{d}{dx}\left[\lambda x - \exp(-x) \right] = \lambda + e^{-x} = 0$$

$$\lambda = - e^{-x} \quad | \log$$

$$x = -\log(-\lambda)$$

① Obtain optimal intercept as a function of λ
- For all x, $g(\lambda)$ is optimal for $f(x)$

Substituting x with $-\log(-\lambda)$ in functional for $g(\lambda)$ gives optimal $g(\lambda)$

$$\Rightarrow g(\lambda) = \lambda(-\log(-\lambda)) - e^{\log(-\lambda)}$$

$$= -\lambda \log(-\lambda) + \lambda$$

$$= \lambda - \lambda \log(-\lambda)$$

POINT: Same component of previously obtained optimal approximation

$$g(x, \lambda) = \lambda x \underbrace{- \lambda + \lambda \log(-\lambda)}_{-g(\lambda)}$$

For concave functions

Switch max → min

② For a particular x (?) find optimal value of variational param λ

$$\boxed{\begin{aligned} f(x) &= \min_\lambda \left[\lambda x - g(\lambda) \right] \\ g(\lambda) &= \min_x \left[\lambda x - f(x) \right] \end{aligned}}$$

① Find variational

function $g(\lambda)$ (?)
OPTIMAL $\forall x$ for
approx. of $f(x)$

Neither convex / concave functions

Like Sigmoid function!

- Need ① $\boxed{\text{invertible transform}}$ → convex form
  - Function or arguments

⇒ Find ② $\boxed{\text{conjugate function } g(\lambda)}$ and transform back

Ex: Upper bound for sigmoid function $\sigma(x) = \dfrac{1}{1+e^{-x}}$

- Transformation : log( ) → Concave function ⇒ Upper bound

$$\log(\sigma(x)) = \underbrace{\log(1)}_{\to 0} - \log(1+e^{-x}) = -\log(1+e^{-x})$$

- Solving $g(\lambda)$:

Function being approximated

⇒ Upper bound for $\sigma()$

$$g(\lambda) = \min_x \left[\lambda x - f(x)\right], \quad f(x) = \log(\sigma(x)) = -\log(1+e^{-x})$$

$$= \min_x \left[\lambda x + \log(1+e^{-x})\right]$$

$$\frac{d}{dx}\left[\lambda x + \log(1+e^{-x})\right] = 0$$

$$\lambda - \frac{1}{e^{x}+1} = 0$$

$$\frac{1}{e^{x}+1} = \lambda$$

$$e^x + 1 = \frac{1}{\lambda}$$

$$e^x = \frac{1}{\lambda} - 1 = \frac{1}{\lambda}(1-\lambda) \qquad | \log()$$

$$x = \log((1-\lambda)/\lambda) = \log(1-\lambda) - \log(\lambda)$$

$$g(\lambda) = \lambda(\log(1-\lambda) - \log(\lambda)) + \log(1 + e^{-(\log(1-\lambda) - \log(\lambda))})$$

$$e^{a-b} = e^a / e^b$$

$$= \lambda \log(1-\lambda) - \lambda \log(\lambda) + \log(1 + e^{\log(\lambda)} / e^{\log(1-\lambda)})$$

$$= \lambda \log(1-\lambda) - \lambda \log(\lambda) + \log(\underline{1 + \lambda/(1-\lambda)})$$

$$1 + \frac{\lambda}{1-\lambda} = \frac{1-\lambda}{1-\lambda} + \frac{\lambda}{1-\lambda} = \frac{1}{1-\lambda}, \quad \log(a/b) = \log a - \log b$$

$$= \lambda \log(1-\lambda) - \lambda \log(\lambda) + \underbrace{\log 1}_{\to 0} - \log(1-\lambda)$$

Conjugate function

*
$$\boxed{g(\lambda) = -\lambda \log(\lambda) - (1-\lambda)\log(1-\lambda)}$$

~ Binary entropy function w. $p(Y=1) = \lambda$

- Upper bound on $\log(\sigma())$:

$$\log(\sigma(x)) \leq \lambda x - g(\lambda) \qquad | e^{()}$$

- Upper bound on $\sigma()$:

* $\boxed{\sigma(x) \leq \exp(\lambda x - g(\lambda))}$

Ex: Lower bound for sigmoid function

$e^a(e^b + e^c) = e^{a+b} + e^{a+c}$

$$\log(\sigma(x)) = -\log(1 + e^{-x}) = -\log(e^{-\frac{1}{2}x}(e^{\frac{1}{2}x} + e^{-\frac{1}{2}x}))$$

$$= -(\log(e^{-\frac{1}{2}x}) + \log(e^{\frac{1}{2}x} + e^{-\frac{1}{2}x}))$$

$$= -(-\tfrac{1}{2}x + \log(e^{\frac{1}{2}x} + e^{-\frac{1}{2}x}))$$

$$= \tfrac{1}{2}x - \log(e^{\frac{1}{2}x} + e^{-\frac{1}{2}x})$$

POINT: Want a ① convex function representing a ② lower bound on $\log(\sigma(x))$

① Convex function $f''(x) > 0 \;\forall x$ in I if $x \to x^2$

$$\log(\sigma(x)) = \tfrac{1}{2}x - \log(e^{\frac{1}{2}x} + e^{-\frac{1}{2}x}) \geq \underbrace{-\log(e^{\frac{1}{2}x} + e^{-\frac{1}{2}x})}_{\text{② Lower bound}}$$

$\Rightarrow$ $f(x) = -\log(e^{\frac{1}{2}x} + e^{-\frac{1}{2}x})$ for variable $x \to x^2$ is a lower bound on $\log(\sigma())$

· Conjugate function

This become the variable as accepted by the function

- lower bound for $\sigma()$

$$g(\lambda) = \max_{x^2}\left[\lambda x^2 - f(\sqrt{x^2})\right]$$

This is the variable optimized for

$$\frac{d}{dx^2}\left[\lambda x^2 - f(\sqrt{x^2})\right] = 0$$ <Confusing part>

Confirmed

$$\Rightarrow \lambda = -\frac{1}{4x}\tanh\left(\frac{x}{2}\right) = -\frac{1}{2x}\left[\sigma(x) - \frac{1}{2}\right]$$

POINT : x is the value corresponding to contact point of tangent line for value of λ of lower bound approximation $z(x,\lambda) = \lambda x - g(\lambda)$ for $\log(\sigma(x))$

Relates approx. $z()$ to local region of $f(x)$

$$\Rightarrow \boxed{x = \xi}$$

POINT : $\lambda \neq \xi$ different meaning

Variational parameter $g(\lambda)$ Optimizes variational $z()$ to $f(x)$

Variational parameter OPTIMAL for $f(x)$ at x

$$\boxed{\lambda(\xi) = -\frac{1}{4\xi}\tanh\left(\frac{1}{2}\xi\right) = -\frac{1}{2\xi}\left[\sigma(\xi) - \frac{1}{2}\right]}$$

- Know x => Know optimal λ value

- Expressing $g()$ in terms of ξ

Knowing optimal λ for an x => Optimal intercept $g()$

$$g(\xi) = \lambda(\xi)\xi^2 + \log\left(e^{\xi/2} + e^{-\xi/2}\right)$$

- Lower bound on $f(x)$ (which is lower bound on $\log(\sigma()))$

$$\log(\sigma(x)) \geq f(x) \geq \lambda x^2 - g(\lambda)$$

$$= \lambda x^2 - \lambda\xi^2 + \log\left(e^{\xi/2} + e^{-\xi/2}\right)$$

$$= \lambda(x^2 - \xi^2) + \log\left(e^{\xi/2} + e^{-\xi/2}\right) \quad | \; e^{()}$$

$$\sigma(x) \geq \exp\left(-\frac{1}{2\xi}\left[\sigma(\xi) - \frac{1}{2}\right](x^2 - \xi^2) + \log\left(e^{\xi/2} + e^{-\xi/2}\right)\right)$$

?

$$\overset{*}{\Rightarrow} \quad \boxed{\sigma(x) \geq \overbrace{\sigma(\xi)\exp\left(\frac{x-\xi}{2} - \lambda(\xi)(x^2-\xi^2)\right)}^{\tau(x,\xi)}}$$

Optimal lower bound of $\sigma(x)$
- Centered at $x = \xi$ (where $\sigma() = \tau()$)

POINT : The regional approx. of $\sigma(x)$ will vary depending on value of ξ

- HOWEVER, the lower bound is OPTIMAL for a given ξ, because the variational parameter λ is optimal for ξ