Assignment 2

Computational Intelligence, SS2020

Team Members		
Last name	First name	Matriculation Number
Blöcher	Christian	01573246
Bürgener	Max	01531577

1 Linear regression

1.1 Derivation of Regularized Linear Regression

1.2 Linear Regression with polynomial features

- Analytical derivation for the Gaussian distribution:

$$p(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{2\sigma^2}}$$

The data is independent identically distributed (iid), therefore the likelihood function is the product of all individual likelihoods

$$P(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) \quad = \quad \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{2\sigma^2}}$$

To convert the product to a sum we apply the natural logarithm.

$$L(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) \quad = \quad ln\left[\prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{2\sigma^2}}\right]$$

$$L(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) \quad = \quad \sum_{n=0}^{N-1}\left[ln(\frac{1}{\sqrt{2\pi\sigma^2}}) - \frac{[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{2\sigma^2}\right]$$

Since we want to find the parameter σ^2, which maximizes the probability of the distance, we derive $L(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p})$ and set it to zero.

$$L(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) \quad = \quad \sum_{n=0}^{N-1}\left[ln(1) - \frac{1}{2}ln(2\pi\sigma^2) - \frac{[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{2\sigma^2}\right] \quad \mid \frac{\partial}{\partial\sigma^2}$$

$$\frac{\partial}{\partial\sigma^2}L(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) \quad = \quad \sum_{n=0}^{N-1}\left[-\frac{1}{\sigma^2} + \frac{[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{\sigma^4}\right] \quad \stackrel{!}{=} 0$$

$$0 \quad = \quad \sum_{n=0}^{N-1}\left[-\frac{1}{\sigma^2} + \frac{[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{\sigma^4}\right]$$

$$\frac{N}{\sigma^2} \quad = \quad \sum_{n=0}^{N-1}\frac{[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{\sigma^4}$$

$$\sigma^2 \quad = \quad \frac{\sum_{n=0}^{N-1}[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,\mathbf{p})]^2}{N}$$

- Analytical derivation for the Exponential distribution:

$$p(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) = \begin{cases} \lambda_i e^{-\lambda_i[\tilde{d}_n(a_i,\mathbf{p})-d(a_i,p)]} & , \tilde{d}_n(a_i,\mathbf{p}) \geq d(a_i,\mathbf{p}) \\ 0 & , \text{else} \end{cases}$$

$$
\begin{aligned}
L(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) &= \sum_{n=0}^{N-1} ln(\lambda_i) - \lambda_i[\tilde{d}_n(a_i,\mathbf{p}) - d(a_i,\mathbf{p})] \quad \mid \tfrac{\partial}{\partial\lambda_i} \\
\tfrac{\partial}{\partial\lambda_i} L(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) &= \tfrac{N}{\lambda_i} - \sum_{n=0}^{N-1} [\tilde{d}_n(a_i,\mathbf{p}) - d(a_i,\mathbf{p})] \quad \stackrel{!}{=} 0 \\
\tfrac{N}{\lambda_i} &= \sum_{n=0}^{N-1} [\tilde{d}_n(a_i,\mathbf{p}) - d(a_i,\mathbf{p})] \\
\lambda_i &= \frac{N}{\sum_{n=0}^{N-1} [\tilde{d}_n(a_i,\mathbf{p}) - d(a_i,\mathbf{p})]} \quad , \tilde{d}_n(a_i,\mathbf{p}) \geq d(a_i,\mathbf{p})
\end{aligned}
$$

1.3 (Bonus) Linear Regression with radial basis functions

2 Logistic Regression

2.1 Derivation of Gradient

- Analytical conversion of the ML estimation equation:

$$
\begin{aligned}
\hat{\mathbf{p}}_{ML}(n) &= \underset{\mathbf{p}}{\text{argmax}} \prod_{i=0}^{N_A-1} p(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) \\
\hat{\mathbf{p}}_{ML}(n) &= \underset{\mathbf{p}}{\text{argmax}}\, ln \left[\prod_{i=0}^{N_A-1} p(\tilde{d}_n(a_i,\mathbf{p}) \mid \mathbf{p}) \right] \\
\hat{\mathbf{p}}_{ML}(n) &= \underset{\mathbf{p}}{\text{argmax}} \sum_{i=0}^{N_A-1} ln \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \right) - \frac{[\tilde{d}_n(a_i,\mathbf{p}) - d(a_i,\mathbf{p})]^2}{2\sigma_i^2}
\end{aligned}
$$

Because in scenario 1 we only use Gaussian models for all anchors that were calibrated with the same distance to the reference position, we can assume that $\sigma_i^2 = \sigma^2\ \forall i$. That means the ln-term can be neglected since it only shifts the value of the maximum by a constant but does not affect its position. Similarly $\frac{1}{2\sigma^2}$ can be omitted, as it is also just a scaling factor. Furthermore $\underset{\mathbf{p}}{\text{argmax}}\,(-\dots)$ is equivalent to $\underset{\mathbf{p}}{\text{argmin}}\,(\dots)$. Thus:

$$
\hat{\mathbf{p}}_{ML}(n) = \underset{\mathbf{p}}{\text{argmin}} \sum_{i=0}^{N_A-1} [\tilde{d}_n(a_i,\mathbf{p}) - d(a_i,\mathbf{p})]^2 = \hat{\mathbf{p}}_{LS}(n)
$$

2.2 Logistic Regression training with gradient descent