Lösningsförslag till tentamen i

Kompression av ljud och bild
TSBK35

24:e mars 2021

1 a) Golombkodning är bra för geometriska (och näraliggande) fördelningar, men textfiler (oberoende av språk) kan inte förväntas ha den typen av fördelning.

b) Eftersom närliggande tecken i en textfil är starkt korrelerade, vill man använda en kodningsmetod som kan utnyttja detta beroende. Därför är LZSS en bättre metod än minnesfri huffmankodning.

2 Se kurslitteraturen.

3 Se kurslitteraturen

4 a) Se kurslitteraturen

b) Se kurslitteraturen

c) Se kurslitteraturen

5 Se kurslitteraturen

6 Lösning för $Q = 0$.

a) Den teoretiskt lägsta datatakten ges av källans entropi:

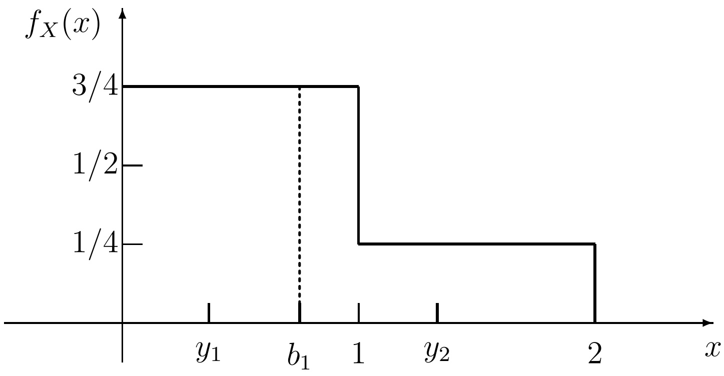
$$H \approx 2.5654 \text{ bitar/symbol}$$

b) Ett exempel på en huffmankod för källan är

symbol	sannolikhet	längd	kodord
a	0.4	1	0
b	0.15	3	100
c	0.1	4	1100
d	0.1	4	1101
e	0.11	3	101
f	0.06	4	1110
g	0.05	5	11110
h	0.03	5	11111

Huffmankoden har en kodordsmedellängd som är 2.62 bitar/kodord, och datatakten 2.62 bitar/symbol.

7 Antag att vi har beslutsgränserna $b_0 = 0, b_1$ och $b_2 = 2$. Rekonstruktionspunkterna är y_1 och y_2. Antag att $b_1 \leq 1$.



De nödvändiga villkoren för minimal distorsion ger att

$$\left.\begin{array}{l} y_1 = \frac{b_1}{2} \\ y_2 - b_1 = b_1 - y_1 \end{array}\right\} \Rightarrow y_2 = \frac{3}{2} b_1 \quad (*)$$

Dessutom ska y_2 placeras i det högra områdets centroid

$$y_2 = \frac{\int_{b_1}^{2} x f_X(x)dx}{\int_{b_1}^{2} f_X(x)dx} = \frac{\frac{3}{4} - \frac{3b_1^2}{8}}{1 - \frac{3b_1}{4}} = \frac{6 - 3b_1^2}{8 - 6b_1}$$

Tillsammans med (*) får vi ekvationen

$$b_1^2 - 2b_1 + 1 = 0$$

som har en dubbelrot i $b_1 = 1$. Alltså ges den optimala kvantiseraren av

$$y_1 = \frac{1}{2}; \quad y_2 = \frac{3}{2}; \quad b_0 = 0; \quad b_1 = 1; \quad b_2 = 2$$

Distorsionen ges av

$$D = \int_0^1 \left(x - \frac{1}{2}\right)^2 \frac{3}{4}\, dx + \int_1^2 \left(x - \frac{3}{2}\right)^2 \frac{1}{4}\, dx = \frac{1}{12}$$

Notera att den optimala kvantiseraren är en likformig kvantiserare, trots att fördelningen inte är likformig.

8 Kvantiseringen är så fin att vi kan göra approximationen att prediktionen görs från originalsignalen.

Vi kan anta att datatakten från den aritmetiska kodaren är samma som entropin för den kvantiserade signalen. Den fina kvantiseringen gör också att prediktionsfelet approximativt är gaussiskt. Då blir distorsionen

$$D \approx \sigma_d^2 \cdot \frac{\pi e}{6} \cdot 2^{-2R}$$

där σ_d^2 är prediktionsfelets varians, och $R = 6$.

$$p_n = a_1 \cdot \hat{Y}_{n-1} + a_2 \cdot \hat{Y}_{n-2} + a_3 \cdot \hat{Y}_{n-3} \approx a_1 \cdot Y_{n-1} + a_2 \cdot Y_{n-2} + a_3 \cdot Y_{n-3}$$

Hitta a_1, a_2 och a_3 som minimerar prediktionsfelets varians σ_d^2.

$$\begin{aligned}
\sigma_d^2 &= E\{(Y_n - p_n)^2\} \approx \\
&\approx E\{(Y_n - a_1 \cdot Y_{n-1} - a_2 \cdot Y_{n-2} - a_3 \cdot Y_{n-3})^2\} = \\
&= (1 + a_1^2 + a_2^2 + a_3^2) R_{YY}(0) - 2a_1 \cdot R_{YY}(1) - 2a_2 \cdot R_{YY}(2) \\
&\quad -2a_3 \cdot R_{YY}(3) + 2a_1 a_2 \cdot R_{YY}(1) + 2a_1 a_3 \cdot R_{YY}(2) + 2a_2 a_3 \cdot R_{YY}(1)
\end{aligned}$$

Derivera med avseende på a_1 och a_2 och sätt lika med 0, vilket ger oss lösningen

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} R_{YY}(0) & R_{YY}(1) & R_{YY}(2) \\ R_{YY}(1) & R_{YY}(0) & R_{YY}(1) \\ R_{YY}(2) & R_{YY}(1) & R_{YY}(0) \end{pmatrix}^{-1} \begin{pmatrix} R_{YY}(1) \\ R_{YY}(2) \\ R_{YY}(3) \end{pmatrix} \approx \begin{pmatrix} 1.4798 \\ -0.7181 \\ -0.1175 \end{pmatrix}$$

$$\sigma_d^2 \approx 0.3485$$

Det resulterande signal-brus-förhållandet blir

$$SNR \approx 10 \cdot \log_{10} \frac{\sigma_Y^2}{\sigma_d^2 \frac{\pi e}{6} 2^{-2R}} \approx 47.19 \text{ [dB]}$$

9 Lösning för $Q = 0$.

Varianser för de fyra transformkomponenterna:

$$\begin{aligned}
\sigma_0^2 &= E\{\theta_0^2\} = \frac{1}{4}E\{(X_0 + X_1 + X_2 + X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0) + 6R_{XX}(1) + 4R_{XX}(2) + 2R_{XX}(3)) \approx 3.6157 \\
\sigma_1^2 &= E\{\theta_1^2\} = \frac{1}{2}E\{(X_0 - X_3)^2\} = \\
&= \frac{1}{2}(2R_{XX}(0) - 2R_{XX}(3)) \approx 0.2213 \\
\sigma_2^2 &= E\{\theta_2^2\} = \frac{1}{4}E\{(X_0 - X_1 - X_2 + X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0) - 2R_{XX}(1) - 4R_{XX}(2) + 2R_{XX}(3)) \approx 0.08294 \\
\sigma_3^2 &= E\{\theta_3^2\} = \frac{1}{2}E\{(-X_1 + X_2)^2\} = \\
&= \frac{1}{2}(2R_{XX}(0) - 2R_{XX}(1)) = 0.08
\end{aligned}$$

Alternativt kan vi beräkna varianserna som diagonalelementen i $\mathbf{A} \cdot \mathbf{R}_X \cdot \mathbf{A}^T$, där

$$\mathbf{R}_X = \begin{pmatrix} 1 & 0.92 & 0.92^2 & 0.92^3 \\ 0.92 & 1 & 0.92 & 0.92^2 \\ 0.92^2 & 0.92 & 1 & 0.92 \\ 0.92^3 & 0.92^2 & 0.92 & 1 \end{pmatrix}$$

Medeldatatakten ska vara 2 bitar/sampel, så vi ska allokera totalt $2 \cdot 4 = 8$ bitar till de fyra transformkomponenterna. Distorsionen minimeras om vi allokerar fyra bitar till θ_0, två bitar till θ_1, en bit till θ_2 och en bit till θ_3. Medeldistorsionen blir

$$D \approx \frac{1}{4}(0.009497 \cdot \sigma_0^2 + 0.1175 \cdot \sigma_1^2 + 0.3634 \cdot \sigma_2^2 + 0.3634 \cdot \sigma_3^2) \approx 0.02989$$

Signal-brus-förhållandet blir

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{1}{D} \approx 15.24 \quad [\text{dB}]$$

Om vi istället använder DWHT så har vi transformmatrisen

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{pmatrix}$$

Vi noterar att θ_0 och θ_2 är exakt samma som för den första transformen. Transformkomponenternas varianser blir

$$\begin{aligned}
\sigma_0^2 &\approx 3.6157 \\
\sigma_1^2 &= E\{\theta_1^2\} = \frac{1}{4}E\{(X_0 + X_1 - X_2 - X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0) + 2R_{XX}(1) - 4R_{XX}(2) - 2R_{XX}(3)) \approx 0.2243 \\
\sigma_2^2 &\approx 0.08294 \\
\sigma_3^2 &= E\{\theta_3^2\} = \frac{1}{4}E\{(X_0 - X_1 + X_2 - X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0) - 6R_{XX}(1) + 4R_{XX}(2) - 2R_{XX}(3)) \approx 0.07706
\end{aligned}$$

Den optimala bitallokeringen blir den samma som för den första transformen, vilket ger oss distorsionen

$$D \approx \frac{1}{4}(0.009497 \cdot \sigma_0^2 + 0.1175 \cdot \sigma_1^2 + 0.3634 \cdot \sigma_2^2 + 0.3634 \cdot \sigma_3^2) \approx 0.02971$$

och signal-brus-förhållandet

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{1}{D} \approx 15.27 \quad [\text{dB}]$$

Dvs, DWHT är bara cirka 0.03 dB bättre.